REMARKS

Claims 20-28 are pending in the application and have been examined. Claims 20-28 stand rejected. Claim 27 has been amended to correct certain formalities not related to patentability. Claims 29-31 have been added in this response and contain no new matter. Applicants respectfully request reconsideration and allowance of Claims 20-31.

The Rejection of Claim 27 Under 35 U.S.C. § 112, Second Paragraph

Claim 27 stands rejected under 35 U.S.C. § 112, second paragraph, for failing to particularly point out and distinctly claim the subject matter of the invention. Applicants have amended Claim 27 to conform with proper Markush terminology in accordance with the Examiner's suggestion. Applicants have also amended Claim 27 by omitting the phrase "(i.e., simple sequence repeats (SSR))" in order to clarify the subject matter of the claimed invention. Applicants respectfully request removal of this ground of rejection.

The Rejection of Claims 20-28 Under 35 U.S.C. § 112, First Paragraph

(Written Description)

Claims 20-28 have been rejected under 35 U.S.C. § 112, first paragraph, as containing subject matter that lacks an adequate written description in the specification. The Examiner has taken the position that the claims contain subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the art that the inventors, at the time the application was filed, had possession of the claimed invention. According to the Examiner, the specification does not provide guidance for the isolation or characterization of DNA from any tree species other than *Pinus taeda*, or for the isolation and characterization of any other type of DNA marker other than SSRs from any tree species other than *Pinus taeda*. The Examiner further notes applicant's failure to provide a conserved nucleotide sequence which

LAW OFFICES OF
CHRISTENSEN O'CONNOR JOHNSON KINDNESS PLLC
1420 Fifth Avenue
Suite 2800
Seattle, Washington 98101
206.682.8100

encompasses molecular markers from a multitude of unrelated tree species. Applicants respectfully disagree for the following reasons.

As an initial matter, applicants wish to point out that the claimed invention is directed to a method of tree breeding using DNA analysis to determine pedigree, and is not claiming specific DNA sequences. As described in the specification, DNA analysis refers to any method of analysis that reveals genotype information. Specification, page 12, lines 15-16. Examples of types of DNA analysis methods (e.g., RFLP, AFLP, etc.) are provided which can be used to identify molecular markers useful in the practice of the invention. Specification, page 12, lines 15-28. Molecular markers, revealed through DNA analysis, are DNA sequence polymorphisms that are used as landmarks to track pedigree. Specification, page 12, lines 4-28. Applicants submit that it is not practicable to disclose every sequence of every marker, and further, disclosure of specific DNA sequences is not required to comply with the written description requirement for the following reasons.

As stated in the Written Description Requirement Guidelines, the review of whether the disclosure satisfies the written description requirement for the claimed subject matter is 1) conducted from the standpoint of one of skill in the art at the time the application was filed, 2) there is an inverse correlation between the level of skill and specificity of disclosure and 3) information which is well known in the art need not be described in detail in the specification. *Fed Reg* Vol. 66, No. 4. Jan 5, 2001 pp. 1099-1107, 1105. The written description requirement is met "if a skilled artisan would have understood the inventor to be in possession of the claimed invention at the time of filing, even if every nuance of the claims is not explicitly described." *Id.* at 1106. Patents and printed publications in the art should be relied upon to determine whether an art is mature and what the level of knowledge and skill is in the art. *Id.*

LAW OFFICES OF
CHRISTENSEN O'CONNOR JOHNSON KINDNESS PLLC
1420 Fifth Avenue
Suite 2800
Seattle, Washington 98101
206.682.8100

Applicants submit Claims 20-28 are supported by an adequate written description because the 1) skill in the art of tree breeding was mature at the time the application was filed and 2) a skilled artisan would have understood based on the specification that the inventors were in possession of the claimed invention.

The specification provides adequate guidance for the isolation of DNA from a multitude of tree species. A working example of DNA isolation is provided in the specification using the commercially available DNeasy 96 DNA extraction kit from Qiagen. Specification, page 22, line 13 to page 14, line 4. One skilled in the art would know that the DNeasy kit is useful to isolate DNA from a wide range of plant species. See, *e.g.*, pages 10-15 of Qiagen's technical bulletin hereto attached as Attachment A.

With respect to methods of DNA analysis for progeny pedigree determination, numerous examples of such methods are provided such as, for example, RFLP analysis, AFLP analysis, RAPD analysis, SSR analysis and so on. See *e.g.*, page 16, lines 8-18, Table 2, and page 26, line 18 to page 31, line 19. Applicants note that the methods of DNA analysis disclosed in the specification were routine, well known in the art, and widely applicable to a variety of tree species (see, e.g., Staub J., *HortScience* 31: 729-739, attached hereto as Attachment B.

Applicants submit therefore, that one skilled in the art of plant breeding would recognize that the applicants had possession of methods of DNA isolation and characterization for a wide variety of tree species at the time the application was filed. Therefore, applicants submit that Claims 20-28 are supported by an adequate written description and respectfully request removal of this ground of rejection.

The Rejection of Claims 20-28 Under 35 U.S.C. § 112, First Paragraph (Enablement)

Claims 20-28 stand rejected under 35 U.S.C. Section 112, first paragraph, for lack of an enabling description in the specification. The Examiner alleges lack of enablement with respect

LAW OFFICES OF
CHRISTENSEN O'CONNOR JOHNSON KINDNESS PLLC
1420 Fifth Avenue
Suite 2800
Seattle, Washington 98101
206.682.8100

to (1) the identification of molecular markers from a variety of tree species, (2) phenotypic determination in a multitude of tree species, (3) the use of polymix breeding coupled with pedigree analysis to select an elite breeding group, and (4) that the practice of the method would require undue experimentation. Applicant respectfully disagrees with the Examiner's conclusions for the following reasons.

The test of enablement is whether one reasonably skilled in the art could make or use the invention from the disclosure in a patent coupled with information known in the art at the time of filing without undue experimentation. Not everything necessary to practice the invention need be disclosed, what is well-known is best omitted. All that is necessary is that one skilled in the art be able to practice the claimed invention, given the level of knowledge and skill in the art. M.P.E.P. Section 2164.08.

A. The Specification Provides Specific Guidance on the Identification of Molecular Markers From a Variety of Tree Species

According to the Examiner, the specification does not provide any guidance for the isolation and incorporation of molecular markers into a polymix-mediated tree-breeding program. Applicants note, as described above, that molecular markers are already present in the genomic DNA of the trees in a breeding program; therefore there is no unpredictable step of incorporating these pre-existing molecular markers into a polymix breeding program. Applicants further submit that the specification provides sufficient guidance in view of the state of the art to enable the claimed invention. As stated by the court in *Enzo Biochem. Inc. v. Calgene*, "[i]t is well settled that patent applications are not required to disclose every species encompassed by their claims, even in an unpredictable art. However, there must be sufficient disclosure, either through illustrative examples or terminology, to teach those of ordinary skill how to make and

LAW OFFICES OF
CHRISTENSEN O'CONNOR JOHNSON KINDNESS PLLC
1420 Fifth Avenue
Suite 2800
Seattle, Washington 98101
206.682.8100

use the invention as broadly as it is claimed." 188 F.3d at 1374, 52 U.S.P.Q.2d at 1138 (Fed. Cir. 1999).

Applicants submit that the specification contains specific guidance on the identification of markers from a variety of tree species. Molecular methods useful in the practice of the present invention allow the determination or inference of an individual's genotype based upon analysis of that individual's chemical constituents. Specification, page 12, lines 4-14. The genotype information is then compared to all potential parent genotype information to infer the pedigree of the individual. As described above, the DNA analysis techniques disclosed in the specification were well known in the art at the time of filing (see, e.g., Staub et al., HortScience 31: 729-740 (1996), attached hereto as Attachment B).

The development and use of informative markers for pedigree analysis in a number of tree species is enabled by the specification in view of the state of the art at time of filing. The specification states that any method of molecular analysis can be used that reveals a sufficient number of genetic polymorphisms (variation in a base pair at a given site within members of the same species) to identify which parental plants are the parents of a particular progeny. Specification, page 14, lines 11-13. Further, an illustrative working example describing use of single nucleotide repeat microsatelles (SSRs) to track parentage in Loblolly pine is provided as a type of marker that is useful in the practice of the method of the invention. See Specification, Example 1, Example 3, Example 4 and Example 5. Specific guidance is provided in the form of primer sequences, including 7 primer pairs useful for analyzing chloroplast microsatellites and 3 primer pairs useful for analyzing nuclear microsatellites. See Table 2. However, applicant's note there is nothing inherently unique about Loblolly pine and the markers chosen to suggest this method is limited to this species. The methods for detecting SSR markers that are described in the specification are equally applicable to the detection of any polymorphic nucleic acid marker

LAW OFFICES OF
CHRISTENSEN O'CONNOR JOHNSON KINDNESS PLLC
1420 Fifth Avenue
Suite 2800
Seattle, Washington 98101
206.682.8100

(i.e., nucleic acid amplification and assessment of distribution, inheritance, and variability of polymorphic markers). Specification, page 26, line 18 to page 31, line 19. Applicants wish to point out that the primer pairs for chloroplast provided in the specification (SEQ ID NO: 1-14), originally described by Vendramin et al, Molecular Ecology, 5: 595-598 (1996), are known to work across a wide range of species due to the high degree of sequence conservation of the chloroplast genome. As stated in the specification, all of the 20 microsatellite primer pairs (Table 2, Example 3) used to amplify simple sequence repeat regions in the chloroplast genome of *Pinus thunbergii* were found to also amplify similar size DNA fragments in *P. taeda.* Specification, page 27, lines 20-24. As further evidence of the broad applicability of the disclosed primers, a review by Anzidei et al. states that the same set of primers described by Vendramin "have been used with success in 110 different conifer species belonging to different taxonomic classifications, in particular to the Pinaceae, Cupressaceae and Taxodiaceae." Anzidei M. et al., In: *European Union DGXII Biotechnology FW IV Research Programme Molecular Tools for Biodiversity*, Gillet. E.M. ed. (1999), attached hereto as Attachment C.

Moreover, applicants submit that numerous DNA markers in addition to single nucleotide repeat microsatelles were well known in the art at the time of filing for many tree species, including for example, chloroplast DNA markers in Douglas fir (see e.g., M.U. Stoehr et al., cited by the Examiner in this Office Action); amplified fragment length polymorphism (AFLP) markers in *Populus spp.* (see e.g., Cervera et al., *Plant Growth Regulation*, 20: 47-52 (1996), attached hereto as Attachment D); microsatellite markers in *Magnolia obovata* (see e.g., Isagi et al., *Heredity 84*: 143-151 (2000), attached hereto as Attachment E), AFLP markers in *Persoonia mollis* (Proteaceae) (see e.g., Krauss and Peakall, *Aust. J. Bot, 46*: 533-546 (1998), attached hereto as Attachment F), and RAPD markers in walnut (*Juglans regis L*) (see e.g., F.P. Nicese et al., *Euphytica 101*: 199-206 (1998), attached hereto as Attachment G).

LAW OFFICES OF
CHRISTENSEN O'CONNOR JOHNSON KINDNESS PLLC
1420 Fifth Avenue
Suite 2800
Seattle, Washington 98101
206.682.8100

Therefore, for the reasons described above, applicants submit that the enablement requirement of pending Claims 20-28 is met in view of the disclosure in the specification, coupled with the knowledge in the art at the time of filing. Applicants respectfully request withdrawal of this ground of rejection.

B. The Specification Provides Specific Guidance on Phenotype Determination in a Multitude of Tree Species

According to the Examiner, the specification does not provide adequate guidance on phenotype determination in a multitude of tree species. As an initial matter, applicants wish to distinguish marker-assisted selection from the method of the invention, which only uses molecular markers to determine the pedigree of selected progeny. Applicants submit that the specification provides sufficient guidance in view of the state of the art to enable the method of evaluating progeny trees using objective criteria to obtain a phenotypic score as claimed. The term "phenotype score" is described in the specification as the objective measurement of any phenotypic trait or characteristic that is desirable in a plant breeding program, such as, for example, disease resistance, growth rate, growth habit, chemical composition of any plant tissue, drought resistance, temperature hardiness, elevation adaptation, fecundity and breeding value. See e.g., Specification, page 10, lines 23-26. The term "objective criteria" is described as the measurement of any plant characteristic or phenotype with any detection or measurement device that provides statistically meaningful data regarding the characteristic or phenotype being measured. See e.g., Specification, page 10, lines 27-32. In addition, methods are provided for statistical analysis of breeding values and heritability determinations. See, e.g., Specification, page 13, lines 20-30.

Further, a working example is provided describing the measurement of exemplary phenotypic traits including height growth, stem diameter growth, straightness, disease resistance,

LAW OFFICES OF
CHRISTENSEN O'CONNOR JOHNSON KINDNESS PLLC
1420 Fifth Avenue
Suite 2800
Seattle, Washington 98101
206.682.8100

insect resistance, general health and deformities. Specification, page 21, lines 14-20. The growth data was analyzed using a best linear unbiased prediction software called GAREML (Dr. Dudley Huber, University of Florida, Gainesville, Florida) that generated breeding values for growth rate for the maternal parent and for every individual progeny. *Id.* Therefore, applicants submit that the specification provides adequate guidance on phenotype determination in a multitude of tree species.

C. The Specification Provides Specific Guidance on the Use of Polymix Breeding Coupled With Pedigree Analysis to Select an Elite Breeding Group

The Examiner takes the view that the specification provides no guidance regarding the use of the claimed method to select elite genotypes, or the use of the method in any tree species other than *Pinus taeda*. Applicants disagree with the Examiner's conclusions and submit that the specification provides specific guidance for selecting elite trees from the progeny. For example, candidate plants are identified from the progeny plants based on having at least one phenotypic characteristic that is statistically better, based upon objective criteria than other progeny plants. Specification, page 13, lines 31-35. The pedigree of the progeny plants is determined using DNA analysis and an elite breeding group is chosen based on high phenotypic scores and low levels of offspring relatedness. In some embodiments, elite plants are selected from the progeny plants based upon a characteristic selected from the group consisting of phenotype score, estimated breeding value, paternal breeding value, maternal breeding value and any combination thereof. See Specification at page 9, lines 27-30. For example, an elite plant that has a high phenotype score and has parents that are of high breeding value is particularly valuable as a breeding parent in the next generation. Knowledge of an elite plant's pedigree allows selection of the next generation of parental plants to maximize the genetic diversity of new breeding groups. See Specification page 16, lines 19-26. Therefore, applicants submit that the

LAW OFFICES OF
CHRISTENSEN O'CONNOR JOHNSON KINDNESS PLLC
1420 Fifth Avenue
Suite 2800
Seattle, Washington 98101
206.682.8100

specification provides adequate guidance on the use of polymix breeding coupled with pedigree analysis to select an elite breeding group

D. The Practice of the Method of the Invention Does Not Require Undue Experimentation.

The Examiner maintains the position that polymix-mediated breeding of trees for phenotypic change is unpredictable, citing Lambeth et al., *Theor Appl Genet* (2001) 103: 930-943. The Examiner also relies on *In re Marzocchi and Horton*, 169 U.S.P.Q. 367 (CCPA 1971) at page 370 for the proposition that applicant's mere assertions in the response or in their own publication are not deemed probative to refute the evidence provided by the Examiner in the form of scientific reasoning and published scientific literature.

Applicants respectfully disagree with the Examiner's interpretation of the *In re Marzocchi* case. The cited case does not support the Examiner's assertion and instead states: "[a]s a matter of Patent Office practice, then, a specification disclosure which contains a teaching of the manner and process of making and using the invention in terms which correspond in scope to those used in describing and defining the subject matter sought to be patented *must* be taken as in compliance with the enabling requirement of the first paragraph of Section 112 *unless* there is reason to doubt the objective truth of the statements contained therein which must be relied on for enabling support 169 USPQ at 369 (*emphasis added*).

With respect to what constitutes undue experimentation, the following factors are relevant: the breadth of the claims, the nature of the invention, the state of the prior art, the relative skill of those in the art, the predictability of the art, the amount of guidance presented, the presence of working examples and the quantity of experimentation necessary MPEP 2164.01(a), *citing In re Wands*, 858 F.2d 731, 737, 8 U.S.P.Q.2d 1400, 1404 (Fed. Cir. 1988).

LAW OFFICES OF
CHRISTENSEN O'CONNOR JOHNSON KINDNESS PLLC
1420 Fifth Avenue
Suite 2800
Seattle, Washington 98101
206.682.8100

Applicants submit the claimed method does not require undue experimentation for the following reasons.

The Examiner relies on White (1996) in *Proc. QFRI-IUFRO Conf. Tree Improvement for Sustainable Tropical Forestry*, for the proposition that the use of polymix breeding may be confounded by the unequal reproductive success of many parents' pollen, which would lead to incorrect measurements of general recombining ability. The White reference at page 133, col 1, second paragraph, refers to potential issues for open-pollination related to tropical species of trees compared to control pollination, and does not state that polymix breeding is unpredictable. Applicants submit that equal fertilization success of the pollen parents of the premix is not essential to the success of the methods of the invention. Moreover, the claimed method actually serves to overcome any effect of unequal pollination through pedigree analysis, thereby allowing for more predictable breeding programs than polymix provides alone. See Specification, page 7, line 34 to page 8, line 12.

The Examiner cites Lambeth et al., page 936 as teaching that the claimed process is unpredictable. Applicants disagree with the Examiner's conclusions for the following reasons. The cited reference describes 3 cases of a total of 45 in which observed genotypes were not consistent with expected genotypes. Applicants submit that these three cases do not render the method unpredictable. The claimed method recites determining the pedigree of a plurality of progeny trees and does not require pedigree assignment of every progeny. Further, as noted in the cited article, the small number of inconsistent cases could be attributed to mislabeled material or incorrect genotyping, are likely not related to the method of DNA analysis. Lambeth et al., page 936. Applicants note that even if cases do arise with inconsistent genotypes, the method provides a selection process wherein they may be simply discarded as not fit for inclusion in the elite breeding group. Further, the method provides additional safeguards such as the routine use

LAW OFFICES OF
CHRISTENSEN O'CONNOR JOHNSON KINDNESS PLLC
1420 Fifth Avenue
Suite 2800
Seattle, Washington 98101
206.682.8100

of markers in confirming parentage in progeny tests, or creating polymixes that avoid mixing pollens known to share the same paternal haplotype. Specification, page 19, line 30 to page 20, line 2.

Therefore, applying the Wands factors to the instant application, it is apparent that a reasonable correlation exists between the scope asserted in the claimed subject matter and the scope of guidance the specification provides because the specification contains adequate, specific guidance on the identification of markers from a variety of tree species and use of the markers in a polymix breeding program. The specification contains a working example using the method of the invention to generate candidates for an elite group, and guidance is provided on the selection of an elite breeding group. Therefore, applicants respectfully submit that the specification provides an enabling disclosure for the claimed invention and request removal of this ground of rejection.

The Rejection of Claims 20-28 Under 35 U.S.C. § 103(a) as Being Unpatentable Over Bridgwater in View of El-Kassaby and Further in View of Stoehr

Claims 20-28 stand rejected under 35 U.S.C. § 103(a) as being unpatentable over Bridgwater (1992) in *Handbook of Quantitative Forest Genetics*, Kluwer Academic Pub., Dordrect, The Netherlands, pages 69-95 in view of El-Kassaby and Ritland (1992) *Theor. Appl. Genet.* 83(6-7):752-8 and Stoehr et al. (1998) *Can. J. For. Res.* 28: 187-95. According to the Examiner, it would have been obvious to one of ordinary skill in the art to utilize the method of polymix tree breeding taught by Bridgwater, and to modify that method by utilizing the pedigree analysis step in the Douglas fir polymix breeding program taught by El-Kassaby and to further modify that method by utilizing the DNA marker taught by Stoehr et al., as suggested by each reference; given the recognition by those of ordinary skill in the art that each would have continued to function in its known and expected manner.

LAW OFFICES OF
CHRISTENSEN O'CONNOR JOHNSON KINDNESS PLLC
1420 Fifth Avenue
Suite 2800
Seattle, Washington 98101
206.682.8100

Applicants submit that the Examiner has failed to establish a *prima facie* case of obviousness. Three requirements are listed in the M.P.E.P. Section 706.02(j) for establishing a *prima facie* case of obviousness. First, there must be some suggestion or motivation, either in the references themselves or in knowledge generally available to one of ordinary skill in the art, to modify the reference or to combine the referenced teachings. Second, there must be a reasonable expectation of success. Finally, the prior art references must teach or suggest all the claim limitations.

For the reasons set forth in detail below, applicants respectfully submit that the burden of establishing a *prima facie* case of obviousness has not been met. First, there is no suggestion to combine or modify the reference's teachings to arrive at the claimed invention. Second, because the cited references teach away from the claimed invention, there can be no reasonable expectation of success for their combined teachings. Applicants remind the Examiner that hindsight and the guidance of the applicant's disclosure cannot be used to reconstruct the claimed invention. The teaching or suggestion to make the claimed combination and the reasonable expectation of success must both be found in the prior art, not in applicant's disclosure. MPEP Section 2143, citing *In re Vaeck*, 947 F.2d 488, 20 U.S.P.Q.2d 1438 (Fed. Cir. 1991). The mere fact that the prior art could be so modified would not have made the modification obvious unless the prior art suggested the desirability of the modification. *In re* Gordon, 221 U.S.P.Q. 1125 (Fed. Cir. 1984). Also, in making a *prima facie* case of obviousness, the teachings of a reference must be taken in its entirety.

The Examiner cites Bridgwater as teaching the advantages of polymix-mediated breeding such as resistance to rust disease, general combining ability and gains in additive genetic variation; wherein one type of polymix scheme is complete nesting involving the use of all pollen parents as females; and wherein the scheme generally costs less than other breeding

LAW OFFICES OF
CHRISTENSEN O'CONNOR JOHNSON KINDNESS PLLC
1420 Fifth Avenue
Suite 2800
Seattle, Washington 98101
206.682.8100

schemes such as diallel crossing. Applicants submit, however, that Bridgwater does not teach or suggest the use of pedigree analysis of progeny or DNA analysis to determine pedigree in a polymix breeding program, as required in the claimed invention. Moreover, applicants submit Bridgwater teaches away from using polymix in a wide variety of tree species due to lack of control of male pedigree based on inbreeding depression which would reduce expected genetic gains. See Bridgwater at page 75.

The Examiner cites El-Kassaby as teaching the use of molecular markers such as isozymes to determine the pedigree of progeny from a polymix cross of Douglas fir trees. Applicants submit El-Kassaby actually teaches away from their invention, for the following reasons. El-Kassaby describes a study using a polymix of three pollen donors chosen based on multilocus allozyme genotypes giving unambiguous determination of paternity to study male reproductive success. The three males in the study showed wide variation reproductive success, leading to the conclusion, in concurrence with Bridgwater, that a drawback of the polymix breeding method is lack of male pedigree control. The solution proposed by El-Kassaby teaches away from the invention by suggesting the use of a polycross with few or single males to determine general combining ability. The references notes "due to the increased co-ancestry among offspring, using fewer males prevents concurrent testing and selecting." El-Kassaby at page 758. This teaching would not motivate one to use polymix for a breeding program with concurrent pedigree testing as claimed. Therefore, applicants submit that the El-Kassaby reference would not provide the required reasonable expectation of success for modifying the method of polymix breeding to include the step of concurrent pedigree analysis of progeny, nor does it teach the use of DNA analysis.

The Examiner cites Stoehr et al. as teaching the use of DNA markers to identify pedigree in Douglas fir, wherein the technique has many advantages including increased accuracy and

LAW OFFICES OF
CHRISTENSEN O'CONNOR JOHNSON KINDNESS^{PLLC}
1420 Fifth Avenue
Suite 2800
Seattle, Washington 98101
206.682.8100

resolution over other markers such as isozymes. However, Stoehr et al. neither suggests nor provides any motivation for using a pedigree and a phenotype score to identify elite trees for use in a next generation of tree breeding, as required by the claimed invention. Rather, Stoehr et al. used a polymorphic genome marker to estimate the level of outside-orchard pollen contamination, supplemental mass pollination efficacies and natural selfing in Douglas fir.

For the reasons noted above, the cited references fail to teach, suggest, provide any motivation to make or otherwise render obvious the claimed invention. Accordingly, applicants respectfully request withdrawal of this ground of rejection.

New Claims 29-31

New Claims 29-31 depend from Claim 20 and further define the tree breeding method of the invention. Claim 29 depends from Claim 27 and recites the limitation that the DNA analysis is performed using single nucleotide analysis. Support for Claim 30 can be found, for example, in the Specification at page 21, line 6 to page 31, line 9. Claim 30 depends from Claim 20 and recites the further limitation that the breeding group consists of conifer species. Support for Claim 30 can be found, for example, in the Specification at page 12, line 29 to page 13, line 2, page 32, lines 2-5 and page 27, lines 15-19. Claim 31 depends from Claim 30 and recites the additional limitation that the DNA analysis is performed using single nucleotide analysis. Support for Claim 31 can be found, for example, in the Specification at page 21, line 6 to page 31, line 9. No new matter has been introduced.

LAW OFFICES OF
CHRISTENSEN O'CONNOR JOHNSON KINDNESS PLLC
1420 Fifth Avenue
Suite 2800
Seattle, Washington 98101
206.682.8100

CONCLUSION

In view of the above amendments and the foregoing remarks, applicants respectfully submit that all the pending claims are in condition for allowance. If any issues remain that may be expeditiously addressed in a telephone interview, the Examiner is encouraged to telephone applicant's attorney.

Respectfully submitted,

CHRISTENSEN O'CONNOR
JOHNSON KINDNESS PLLC

Barry F. McGurl
Registration No. 43,340
Direct Dial No. 206.695.1775

I hereby certify that this correspondence is being deposited with the U.S. Postal Service in a sealed envelope as first class mail with postage thereon fully prepaid and addressed to Mail Stop RCE, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450, on the below date.

Date: 1/27/04 

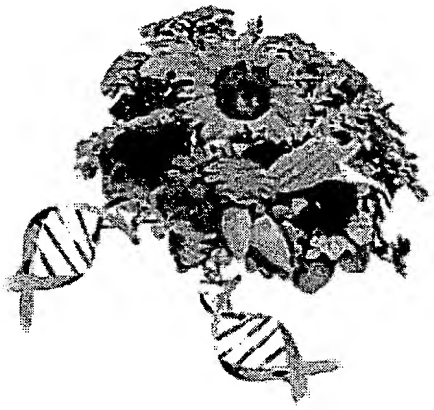
BFM/TJQ:jlj

LAW OFFICES OF
CHRISTENSEN O'CONNOR JOHNSON KINDNESS PLLC
1420 Fifth Avenue
Suite 2800
Seattle, Washington 98101
206.682.8100

Plant Nucleic Acid Purification

Technical hints and applications

QIAGEN



Trademarks and disclaimers

Patented or patent-pending technology and/or registered or registration-pending trademark of QIAGEN: QIAGEN®, QIAshredder™, AirPore™, DNeasy®, HotStarTaq™, Oligotex®, Omniscript™, RNeasy®, and Sensiscript™.

TaqMan is a registered trademark of Roche Molecular Systems, Inc.

Registered names, trademarks, etc. used in this document, even when not specifically marked as such, are not to be considered unprotected by law.

Roundup is a registered trademark of Monsanto Company.

Purchase of products containing QIAGEN Taq DNA Polymerase or HotStarTaq DNA Polymerase is accompanied by a limited license to use them in the Polymerase Chain Reaction (PCR) process for research and development activities in conjunction with a thermal cycler whose use in the automated performance of the PCR process is covered by the up-front license fee, either by payment to Perkin-Elmer or as purchased, i.e. an authorized thermal cycler.

The PCR process is covered by U.S. Patents 4,683,195 and 4,683,202 and foreign equivalents owned by Hoffmann-La Roche AG.

Contents

Effect of DNA Quality on Spectrophotometry

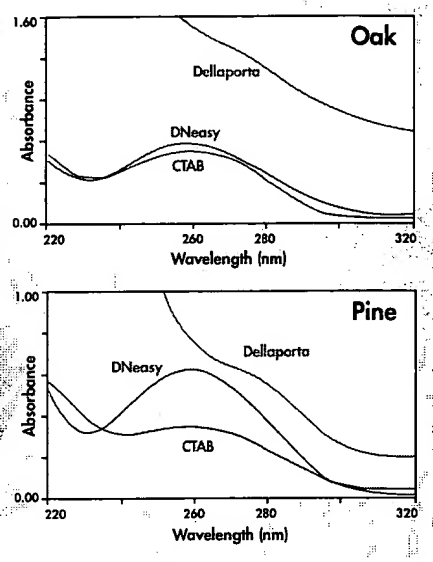


Figure 1. Spectrophotometric scans (220–320 nm) of DNA isolated from leaves/needles using the method of Dellaporta, CTAB, or the DNeasy Plant Mini Kit. Typically, pure DNA shows a symmetrical peak at 260 nm and a smooth profile. Polysaccharides and other secondary metabolites, often copurified with plant DNA isolated using traditional methods, can interfere with spectrophotometric readings (A_{260}/A_{280}), leading to errors in determination of concentration and purity.

Introduction

Recent years have seen an explosion in the number and variety of plant molecular biology applications being used in research laboratories. The isolation of pure nucleic acids from plant materials presents special challenges, and commonly used molecular biology techniques often require adaptation before they can be used with plant samples.

Several plant metabolites have chemical properties similar to those of nucleic acids, making contaminating metabolites difficult to remove from nucleic acid preparations. Co-purified metabolites and contaminants introduced by the purification procedure, such as salt or phenol, can cause inconsistent results in downstream applications. Listed below are some of the most common problems associated with contamination of nucleic acids prepared from plants.

- Inhibition of enzymatic reactions (e.g., restriction digestion, reverse transcription, PCR amplification)
- Inaccurate UV spectrophotometric quantitation (Figure 1)
- Altered electrophoretic mobility (Figure 2)
- Pipetting errors due to increased viscosity
- Nucleic acid degradation during storage

This guide gives an overview of the techniques used for plant nucleic acid purification and provides useful guidelines for successful results.

Effect of DNA Quality on Electrophoretic Mobility

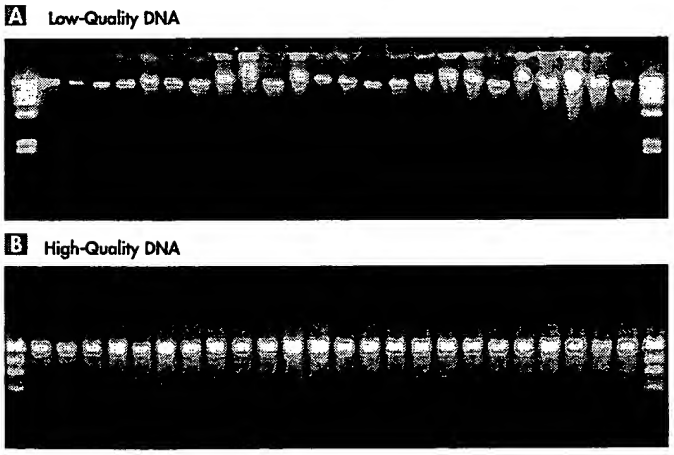


Figure 2. Agarose gel (0.8% TBE) analysis of genomic DNA isolated from plant leaves. A: Low-quality DNA containing residual impurites, which hinder migration of DNA out of wells and cause non-uniform eletrophoretic mobility. B: High-quality DNA, which shows uniform electrophoretic mobility. Left and right end lanes are markers.

Plant starting materials

Isolating intact, pure nucleic acid from plant cells presents special challenges for researchers who study plants. The methods used for growing, harvesting, storing, and preparing plant tissues can influence subsequent nucleic acid purification.

Often, the experimental design dictates which methods are used to prepare plant materials. For example, gene expression analysis may involve specific growth conditions or plants of different ages. DNA may be required from herbarium or forensic samples. However, several parameters, including growth conditions, harvesting method, storage, and tissue disruption technique, can be adjusted when possible to make subsequent isolation easier.

This chapter provides information and general guidelines which may be helpful to optimize conditions for successful nucleic acid purification.

Effect of growth conditions on nucleic acid isolation

Growth conditions influence the production and accumulation of plant metabolites such as polysaccharides, polyphenolics, and flavones. The efficiency of many nucleic acid isolation techniques is affected by the presence of plant metabolites, and the presence of these compounds in RNA and DNA preparations can reduce performance in downstream applications (Figure 3). Many nucleic acid isolation methods recommend growing plants under conditions which do not induce high-level accumulation of plant metabolites.

Metabolite production and accumulation may be affected by several factors:

- Stress (e.g., induced by wounding, desiccation, pathogen infection, or nutrient deficiency)
- Light intensity, spectrum, and duration
- Plant age

Because of the great variation between plants, it is very difficult to make general statements about the production and accumulation of plant metabolites. However, as a general guideline, it is recommended to use healthy, young tissues when possible. In addition, many protocols for "home-made" methods recommend growing plants in darkness for 1 to 2 days before harvesting to prevent high-level accumulation of plant metabolites.

Effect of Carbohydrate Contamination on PCR Performance

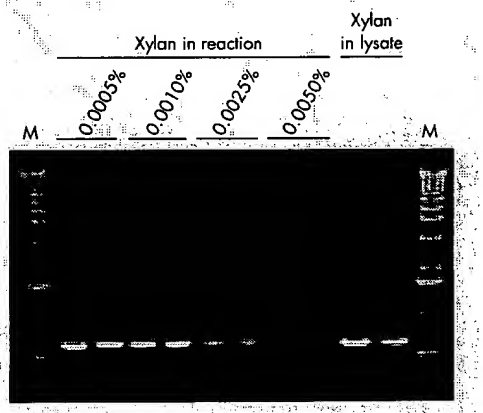


Figure 3. DNA was isolated from wheat leaves using the DNeasy Plant Mini Kit. **Reaction:** Amplification reactions were prepared using 40 ng DNA, 1 unit QIAGEN Taq DNA Polymerase, the indicated concentrations of xylan, and universal primers for the noncoding intergenic spacer region between the chloroplast tRNA genes trnL (UAA) 5' exon and trnL (UAA) 3' exon of chloroplast DNA (reference 3). **Lysate:** xylan was added to wheat lysates before purification so that the xylan concentration in the eluate would be 2.5% if it were not removed during the DNeasy Plant isolation procedure. Amplification reactions were prepared as above. **M:** markers.

Harvesting tissues

The content of nucleic acid and other plant metabolites can vary widely between plant species as well as between organs of the same plant or plants of different ages. Many tissue characteristics affect the ease, efficiency, and yield of nucleic acid purification. Examples of such properties include gene-expression levels, ploidy, vacuolization, and cell wall properties such as lignin content (for an example, see "Nucleic acid content of plant tissues", page 10).

When possible, it is preferable to collect young material (e.g., expanding leaves or needles). Nucleic acid yields from young tissues are often higher than from old tissue, because young tissue generally contains more cells than the same amount of older tissue. In addition, young tissue of the same weight contains fewer metabolites which can affect the performance of downstream applications if not completely removed during nucleic acid purification (see Figure 3, page 5).

When using fresh leaves for mini-preparation, tissue can be harvested by cutting discs (e.g., with a hole puncher) and collecting the disks in the lid of a microcentrifuge tube. A leaf disk with a 1.5 cm diameter weighs 25–75 mg.

Storage of harvested tissues

Tissue damage can result in degradation of nucleic acids. Since tissue can rarely be processed immediately after harvesting, storage conditions that preserve the integrity of the nucleic acids contained in the sample are essential. Improper storage is particularly damaging to RNA, although it can also influence DNA quality.

When DNA is to be isolated, leaves and needles from most species can be stored for up to 24 hours at 4°C without affecting yield or quality. In general, samples that are to be stored for longer than 24 hours should be frozen and kept at –80°C. However, some samples, for example, tree buds, can be stored for several days at 4°C. Tissues stored at 4°C should be kept in a closed container to prevent dehydration. Large samples (e.g., branches) can be stored in a plastic bag containing a wet paper towel.

For RNA isolation, plant material should be frozen in liquid nitrogen immediately after harvesting. Frozen samples can be stored at –80°C indefinitely for later processing. For convenience and efficient use of space, frozen tissue can be disrupted under liquid nitrogen (see "Disruption of plant materials", page 7) and the resulting powder stored at –80°C.

When it is not practical to store frozen samples for DNA preparation, a number of methods are available for drying plant tissue using, for example, silica gel, food dehydrators, or lyophilizers (1). To prevent DNA degradation, material should be completely desiccated in less than 24 hours. Dried samples should be kept in darkness at room temperature under desiccating or hermetic conditions for long-term storage (2). Depending on how the material was handled, the DNA in herbarium and forensic samples may be degraded. When such samples are to be used in PCR, primers should be designed to amplify short (<500 bp) segments.

Disrupted plant material can also be stored in DNeasy® lysis buffer AP1 at room temperature for several months without appreciable DNA degradation (Figure 4).

Safe Sample Storage using Buffer AP1

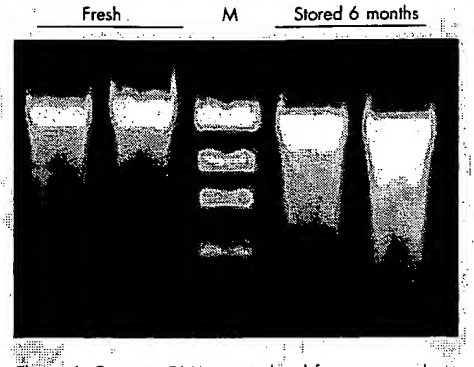


Figure 4. Genomic DNA was isolated from young wheat leaves using the DNeasy Plant Mini Kit. After disruption, lysis buffer AP1 was added to the tissue and the mixture was incubated at 65°C for 10 min. DNA isolation was carried out according to the protocol either immediately after lysis **(Fresh)** or after 6 months storage at room temperature **(Stored 6 month)**. Eluates were run on a 0.8% TBE agarose gel. **M:** markers.

Disruption of plant material

Complete disruption of cell walls, plasma membranes, and organelle membranes is essential to release all the nucleic acids contained in tissue. Insufficient disruption of starting material will lead to low yield. Cell wall properties vary widely between different species and different methods are required to achieve complete disruption.

Disruption using mortar and pestle

The most common disruption method involves freezing samples in liquid nitrogen and grinding with a mortar and pestle (Table 1).

Table 1. Typical protocol for disrupting plant samples using mortar and pestle

1. Freeze tissue in liquid nitrogen immediately after harvesting. Do not let the sample to thaw at any time during disruption.
2. Precool mortar to –20°C and keep on dry ice.
3. Pour liquid nitrogen into the mortar, and precool pestle by placing the grinding end in the liquid nitrogen.
4. Place frozen tissue in mortar and grind until a fine, whitish powder results.
5. Add liquid nitrogen as necessary, being careful the sample does not spill out of the mortar.
6. Using a precooled spatula, transfer the powder to pre-cooled containers of the appropriate size. To avoid thawing, large samples may be transferred to several containers.
7. Ensure all liquid nitrogen has evaporated before closing the container. To prevent the sample from thawing after evaporation, the container should be cooled by placing it in dry ice or liquid nitrogen.

Several modified protocols using liquid nitrogen and small containers, such as microcentrifuge tubes, are available for disruption of small tissue samples. Most protocols are adapted for a specific tissue (e.g., young leaves) and include crushing tissue with a glass rod, plastic pestle, or wooden stick. These modified methods may result in DNA yields which are 20–80% of the yields obtained using standard disruption using a mortar and pestle. A side-by-side comparison of modified and standard methods is recommended before using a modified method for tissue disruption. Modified disruption methods are not recommended for RNA isolation, as variation in yield prevents accurate quantitative analysis.

Tissue powder can be used directly for nucleic acid purification. After disruption and lysis, the lysate may be viscous and must be homogenized (see page 9). Homogenization is particularly important for RNA isolation.

Disruption using a rotor–stator homogenator

Rotor–stator homogenizers disrupt relatively soft plant tissues in the presence of lysis buffer. The rotor turns at a very high speed causing the sample to be disrupted and homogenized by a combination of turbulence and mechanical shearing. These homogenizers may not disrupt tough tissue, such as roots, and their use is not universally recommended.

Rotor–stator homogenizers are available in different sizes and operate with differently sized probes. Probes with diameters of 5–7 mm are suitable for volumes up to 300 µl and can be used for homogenization in microcentrifuge tubes. Probes with diameters of ≥10 mm require larger tubes.

When disrupting tissue with a rotor–stator homogenator, foaming should be kept to a minimum by using vessels of the appropriate size and by keeping the tip of the homogenizer submerged.

Disruption using a mixer mill

Mixer milling disrupts cells and tissues by rapid agitation in the presence of beads made of tungsten carbide, steel, or glass. Disruption is caused by the shearing and crushing action of the beads as they collide with the cells. When using fresh leaf tissue, most samples can be disrupted in the presence of lysis buffer. Alternatively, disruption of frozen plant material can be performed without lysis buffer if the beads and disruption vessel are precooled with liquid nitrogen. Samples should be disrupted in the presence of either lysis buffer or liquid nitrogen to preserve the quality of the contained nucleic acids.

Optimal disruption parameters must be determined empirically for each application. Disruption efficiency is influenced by the following factors.

- Size and type of bead
- Speed and configuration of agitator
- Duration of disruption
- Amount of starting material

The Mixer Mill MM 300 is the first commercially available high-throughput system designed for simultaneous, rapid, and effective disruption of plant samples (Figure 5). This mixer mill allows up to 192 samples to be disrupted in just 2–4 minutes, and is ideal for use with high-throughput purification formats such as the DNeasy 96 Plant Kit.

Recommendations for disruption and homogenization

In our hands, similar DNA yields were obtained using either a mortar and pestle or mixer mill. Disruption efficiency using other methods depends on the starting material (e.g., tissue type, plant age), and yields may be 20–80% of those obtained using a mortar and pestle or mixer mill. Incomplete disruption always results in reduced yields.

When preparing plant nucleic acids, samples disrupted by grinding in a mortar and pestle must be further homogenized to reduce viscosity caused by high-molecular-weight cellular components such as complex carbohydrates.

The QIAshredder™ unit has been designed for efficient, cross-contamination–free homogenization of cell and tissue lysates. Lysate is loaded onto the QIAshredder spin column which is placed in a collection tube. After spinning in a centrifuge, the homogenized lysate is collected. Use of the QIAshredder unit also improves nucleic acid purification by removing cell debris and precipitates from cleared lysates, and is included in all DNeasy Plant and RNeasy® Plant kits.

Efficient High-Throughput Disruption of Plant Samples

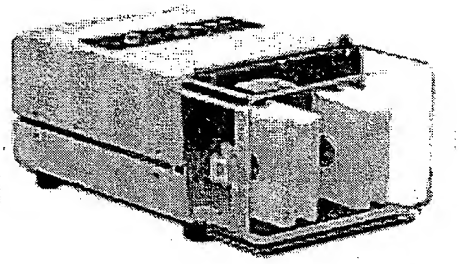


Figure 5. The Mixer Mill MM 300 allows processing of up to 192 plant samples in just 2–4 minutes.

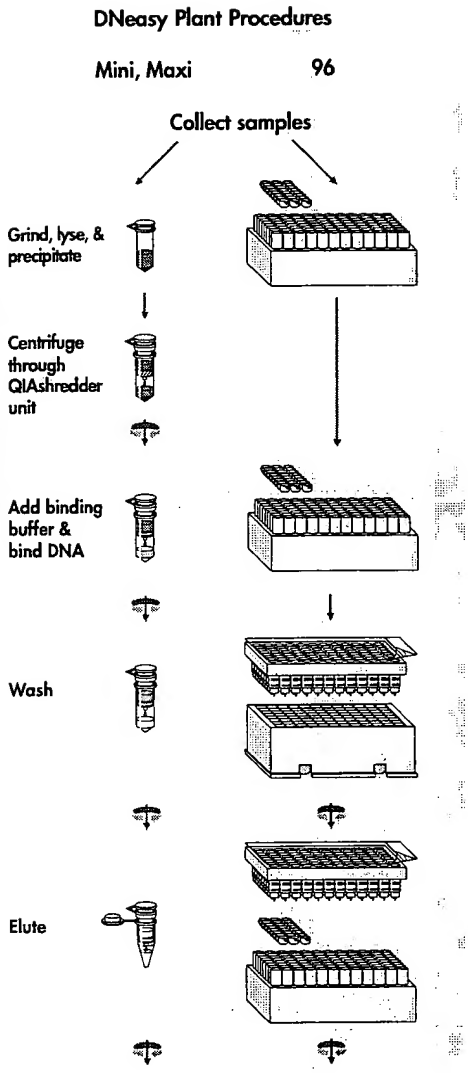


Figure 6. The DNeasy Mini, Maxi, and 96 Plant procedures.

Nucleic acid content of plant tissues

The nucleic acid content can vary widely between different plant starting materials. For example, a tissue sample comprised of small cells will have a higher cell density, and therefore is likely to contain more nucleic acids than a sample of the same size which is comprised of larger cells. In addition, DNA contents depend on the haploid genome size and the ploidy of the sample. Arabidopsis has a small diploid genome and correspondingly lower DNA yields than wheat which has a large hexaploid genome (see Table 2). RNA content varies less predictably than DNA content. Highly proliferating tissues, such as meristems, typically contain more RNA than mature tissues. This variation in nucleic acid content should be considered when purifying nucleic acids.

Table 2. Typical DNA yields from arabidopsis and wheat

Plant	Genome size	Ploidy	Typical yield from 100 mg fresh tissue
Arabidopsis	1.9×10^8 bp	Diploid	3 µg
Wheat	1.7×10^{10} bp	Hexaploid	30 µg

DNA isolation methods

Plant molecular biology studies often require a simple, rapid, and reproducible method for preparing DNA from a wide variety of species. A number of factors that can affect the yield and purity of DNA must be considered, including incomplete cell lysis and carryover contamination of carbohydrates, polyphenolics, flavones, and other metabolites. DNA isolation methods must be scaleable for different sample sizes and provide sufficient throughput to meet demanding project timelines.

DNeasy® Plant kits

With the DNeasy Plant procedure (Figure 6), plant cells or tissues are first mechanically disrupted and then lysed by the addition of lysis buffer and incubation at 65°C. During this step, RNase contained in the lysis buffer digests RNA in the lysate. After lysis, proteins and polysaccharides are removed by salt precipitation. Precipitates and cellular debris are removed in a single step by a brief spin through a QIAshredder unit (see page 9).

The cleared lysate is transferred to a new tube and a binding buffer containing ethanol is added to promote binding of the DNA to the DNeasy membrane. The sample is then applied to a DNeasy spin column or a 96-well plate and spun briefly in a centrifuge. Contaminants, such as proteins and polysaccharides, are efficiently removed by two stringent wash steps. The highly specific binding properties of the DNeasy membrane allows efficient purification and eliminates the need for additional extraction or precipitation steps which are often required for traditional isolation methods. Pure DNA is eluted in a small volume of low-salt buffer or water.

DNeasy purified DNA typically has A_{260}/A_{280} ratios of 1.7–1.9, and absorbance scans show a symmetric peak at 260 nm, confirming high purity.

The DNA-binding DNeasy membrane is available in either spin-column or 96-well formats. The combination of easy, high-throughput disruption using the Mixer Mill MM 300 (see page 8) and reliable purification using the DNeasy 96 Plant Kit provides convenient DNA isolation from plant tissues in 96-well format (Figure 7).

DNeasy Plant Kits have been used to isolate high-quality DNA from a wide range of plant species and tissues, including troublesome sources rich in polysaccharides and polyphenolics (Figure 8 and Table 3, next page).

Reproducible DNA Yield and PCR Performance

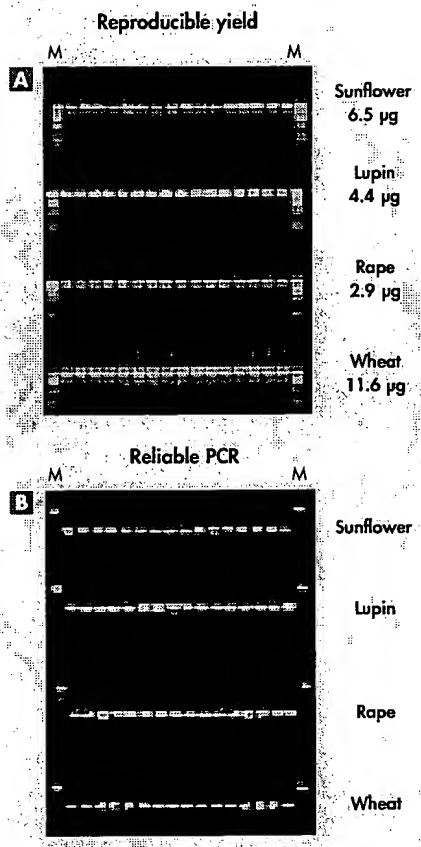


Figure 7. Genomic DNA was purified from different plant species using the DNeasy 96 Plant Kit following disruption in liquid nitrogen. DNA was eluted in 2 x 100 µl Buffer AE. **A:** Ten µl of each eluate was loaded per lane. Average yields are given next to the plant names. **B:** Genomic DNA purified from different plant species was amplified by PCR. Five µl of each eluate was used as template, and 30 µl of each PCR product was loaded in each lane. **M:** markers.

PCR Analysis of DNA from Different Plant Species

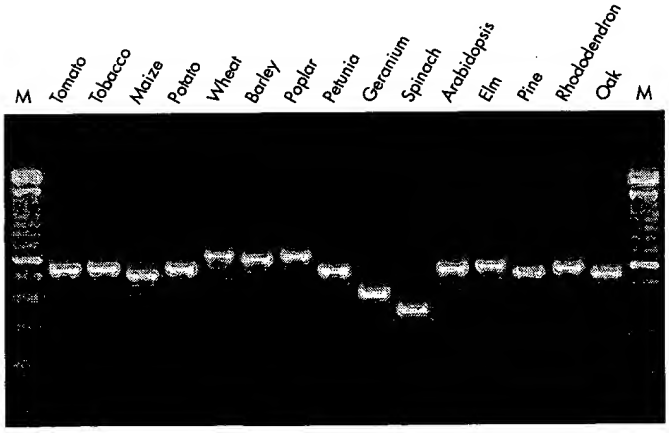


Figure 8. DNA (10 ng) from the indicated leaves or needles was amplified using universal primers for the noncoding intergenic spacer between the tRNA genes trnL (UAA) 5' exon and trnL (UAA) 3' exon of cpDNA (reference 3). **M:** markers.

Table 3. Selection of plant species processed with DNeasy Plant kits

Abies alba (silver fir)	Nicotiana tabacum (tobacco)
Aesculus hippocastanum (horse chestnut)	Oryza sativa (rice)[4]
Arabidopsis thaliana (thale cress)	Pelargonium sp. (geranium)[4]
Avena sp. (oat)	Petunia sp.[4]
Brassica napus (oilseed rape)	Pinus sylvestris (Scotch pine) P. brutia[5]
Brassica oleracea (kohlrabi)	Populus tremula (aspen)
Chicorium endivia (chicory)	Pseudotsuga menziesii (Douglas fir)
Citrullini lanatus (water melon)	Quercus robur Q. petrea (oak)[6,7]
Egeria sp.	Rhododendron sp.[2,4]
Fagus sylvatica (beech)[1]	Rudus idaeus (raspberry)
Helianthus spp. (sunflower)	Solanum tuberosum (potato)
Hordeum vulgare (barley)[2]	Sphagnum palustre (moss)
Humulus sp. (hops)	Spinacia oleracea (spinach)
Hydrilla sp.	Taxus baccata (yew)
Kalanchoe spp.	Triticum aestivum (wheat)[4]
Lupinus sp.	Ulmus glabra (elm)[6]
Lycopersicon esculentum (tomato)[3]	Vitis spp. (grape)[6]
Myriophyllum sp.	Zea mays (maize)

Young leaves or needles (and other tissues, as indicated) were collected and immediately flash frozen. DNA isolation was then performed with the DNeasy Plant Mini Kit. [1]Beechnut, [2]dried leaves, [3]callus, [4]leaves from adult plant, [5]endosperm, [6]old leaves, rich in carbohydrates, [7]buds. For more information on DNA isolation from other species including fungi, call QIAGEN Technical Services or your local distributor.

CTAB lysis

This "home-made" DNA isolation method uses the detergent cetyltrimethylammonium bromide (CTAB) to lyse plant cells (4–6). After lysis, contaminants are removed by a chloroform extraction step. During extraction, it is essential that the correct salt concentration is used to ensure that contaminants are separated into the organic phase and DNA stays in the aqueous phase. DNA is recovered from the aqueous phase with a subsequent precipitation either by adding alcohol or lowering the salt concentration so that DNA forms insoluble complexes with the CTAB. DNA preparations isolated using this method may contain enzyme-inhibiting contaminants and therefore may not be sufficiently pure for sensitive downstream applications such as PCR (Figures 9 and 10).

Effect of Contaminants on PCR Performance

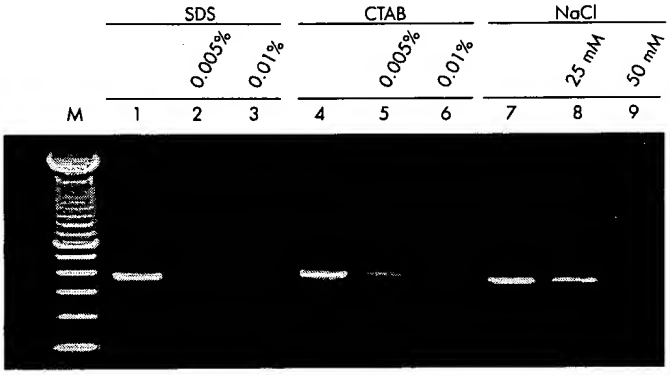


Figure 9. DNA was isolated from oats **(lanes 1–3)**, spinach **(lanes 4–6)**, or kohlrabi **(lanes 7–9)** using the DNeasy Plant Mini Kit. A 400 bp fragment was amplified from 40 ng purified DNA using universal primers for the noncoding intergenic spacer region between the chloroplast tRNA genes trnL (UAA) 5' exon and trnL (UAA) 3' exon of cpDNA (reference 3). Reactions were prepared with the indicated concentrations of SDS, CTAB, or NaCl (contaminants typically found in DNA solutions prepared using the Dellaporta or CTAB methods). M: markers.

Table 4. DNeasy Plant kit specifications

	DNeasy Plant Mini	DNeasy Plant Maxi	DNeasy Plant 96
Amount starting material (maximum)	100 mg wet	1 g wet	50 mg wet
	20 mg dry	200 mg dry	10 mg dry
DNA isolated	Total DNA (genomic, chloroplast, mitochondrial)		
Size of isolated DNA	Up to 40 kb, average 20–25 kb		
DNA binding capacity*	50 µg	500 µg	50 µg
Typical DNA yield†	3–30 µg	30–260 µg	2–12 µg
Elution volume (minimum)	50 µl	500 µl	50 µl
Processing time	<1 h	<2 h	<2 h

* DNA content of most samples does not exceed the binding capacity of the DNeasy membrane.

† Nucleic acid content varies widely between different sources; see "Nucleic acid content of plant tissues", page 10.

Comparison of CTAB and DNeasy Methods: PCR performance

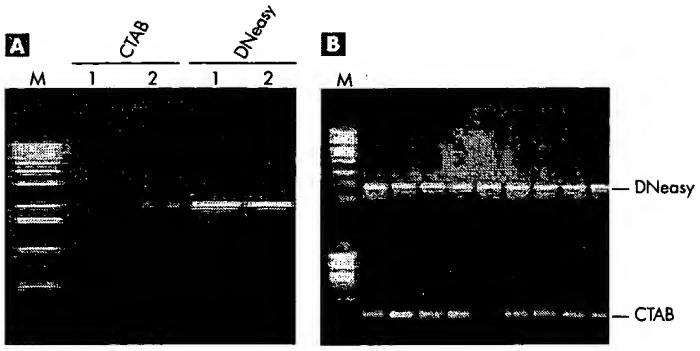


Figure 10.
A: DNA was isolated from arabidopsis leaves using either CTAB lysis (**CTAB**) or the DNeasy Plant Maxi Kit (**DNeasy**). Amplification reactions were prepared using purified DNA (**1**: 50 pg; **2**: 100 pg) and primers to the Akin 10 gene. **M**: markers. (Data kindly provided by Alain Lecharny, Institut de Biotechnologie des Plantes, UMR CNRS-UPS Orsay, France.)
B: DNA was purified from petunia using the DNeasy 96 Plant Kit (**DNeasy**) and a conventional CTAB-based purification method (**CTAB**). The purified DNA was used as template in a PCR to amplify a 600 bp fragment of the glucanase gene. **M**: markers. (Data kindly provided by M. D'Hauw and T. Gerats, Department of Plant Genetics, University of Gent, Gent, Belgium.)

Dellaporta (salting-out) method

This method (7) involves grinding plant tissue in an SDS-containing lysis buffer, filtering the lysate, and precipitating proteins and other compounds contained in the lysate with high salt concentrations. Removal of proteins and other contaminants using this salting-out method may be inefficient. RNase treatment and repeated alcohol precipitation are typically necessary before the DNA can be used in downstream applications. Difficult samples may require manual removal of precipitated DNA from the alcohol suspension to reduce coprecipitation of contaminants by centrifugation. However, this step may not sufficiently remove contaminants, and DNA preparations may contain enzyme-inhibiting compounds. Yields and purity using this method are often variable.

ROSE method

The ROSE method involves disruption and lysis of plant cells followed by incubation at high temperatures (90°C for 20 minutes, 8). The lysate is then used directly in downstream applications. Considered a "quick-and-dirty" technique, this method may not be suitable for extremely sensitive applications because isolated DNA often contains enzyme-inhibiting contaminants (see Table 5, next page). Furthermore, high levels of contamination often result in DNA degradation during storage. Therefore, the ROSE method is appropriate for a limited range of applications.

Table 5. Comparison of DNeasy and ROSE methods: PCR using DNA isolated from gymnosperm species

	Standard PCR		RAPD	
Species (family)	DNeasy	ROSE	DNeasy	ROSE
Cycas circinalias L. (Cycadaceae)	3	3	3	0
Ginkgo biloba L. (Ginkgoaceae)	3	3	3	0
Gnetum gnemon L. (Gnetaceae)	3	3	3	0
Ephedra distachya ssp. helvetica L. (Ephedraceae)	3	2	3	0
Abies alba Mill. (Pinaceae)	3	2	3	0
Cedrus atlantica (Endl.) Manetti ex Carr. (Pinaceae)	3	3	3	0
Larix decidua Mill. (Pinaceae)	3	2	3	0
Picea abies (L.) Karst. (Pinaceae)	3	3	3	2
Pinus sylvestris L. (Pinaceae)	3	3	3	2
Pseudotsuga menziesii (Mirb.) Franco (Pinaceae)	3	2	3	3
Tsuga canadensis (L.) Carr. (Pinaceae)	3	3	3	0
Podocarpus lawerncei Hook.f. (Podocarpaceae)	3	1	3	0
Agathis brownii L.H. Bailey (Araucariacea)	3	0	3	0
Araucaria angustifolia (Bertol.) Kuntze (Araucariacea)	3	0	3	0
Sciadopitys verticillata (Thunb.) Schinz and Zucc. (Scaidopitaceae)	3	3	3	0
Taxus baccata L. (Taxaceae)	3	3	3	0
Torreya nucifera (L.) Schinz and Zucc. (Taxaceae)	3	3	3	3
Cephalotaxus harringtonia var. drupaceae (Forbes) K. Koch (Cephalotaxaceae)	3	3	3	3
Cryptomeria japonica (L.f.) D. Don (Taxodiaceae)	3	0	0	0
Cunninghamia lanceolata (Lamb.) Hook. (Taxodiaceae)	3	0	3	0
Metasequoia glyptostroboides Hu and Cheng (Taxodiaceae)	3	1	3	0
Sequoia sempervirens (D. Don)	3	1	3	0
Sequoiadendron giganteum (Lindl.) Buchh. (Taxodiaceae)	3	2	3	0
Taxodium distichum (L.) A. Rich. (Taxodiaceae)	3	0	3	0
Callitris preissii Miq. (Cupressaceae)	3	3	3	0
Calocedrus decurrens (Torrey) Florin Cupressaceae)	3	3	3	0
Cupressocyparis x leylandii Dallimore and A.B. Jackson (Cupressaceae)	3	2	3	0
Cupressus arizonica Greene (Cupressaceae)	3	3	3	0
Tetraclinis articulata (Vahl) Masters (Cupressaceae)	3	0	3	0
Thuja plicata Donn ex D. Don (Cupressaceae)	3	1	3	0
Thujopsis dolabrata Schinz and Zucc. (Cupressaceae)	3	0	3	0
Waddringtonia cedarbergensis J.A. Marsh (Cupressaceae)	3	0	3	0

DNA was extracted from 32 gymnosperm species using either the DNeasy Plant kit or the ROSE method. Isolated DNA was PCR amplified in triplicate using either primers to an intronic region of the trnL (UUA) chloroplast gene or RAPD primers (reference 9). The table indicates the number of successful reactions. (Data kindly provided by C. Sperison, F. Gugerli, U. Büchler, and G. Mátyás, Biodiversity Department, Swiss Federal Research Institute, WSL, Birmendorf, Switzerland.)

CsCl density gradient

Plant DNA can be isolated by centrifugation through a cesium chloride (CsCl) density gradient (10). After plant cells are lysed with detergent and treated with protease, the cleared lysate is precipitated with isopropanol. The resuspended DNA is then mixed with CsCl and ethidium bromide and centrifuged for several hours. Although this method allows the isolation of high-quality DNA, it is time consuming, labor intensive, and expensive, making it inappropriate for routine use.

Comparison of plant DNA purification methods

Table 6 summarizes several features of the DNA purification methods mentioned in the previous sections.

Table 6. Comparison of plant DNA purification methods

	DNeasy Plant kit	CTAB	Dellaporta	ROSE	CsCl gradient
Sample source	Plant cells and tissues	Plant cells and tissues	Fresh plant tissue	Plant tissues and cells	Plant tissues*
Can be used with a broad range of plant species?	Yes	No†	No†	No†	Yes
DNA quality	High	Medium	Low	Very low	High
Alcohol precipitation required?	No	Yes	Yes	No	Yes
Preparation time for 24 samples	<1 h‡	2–4 h (plus overnight resuspension)	2–4 h (plus overnight resuspension)	1 h	12 h (for 8 samples)§
Reproducibility	High	Variable	Variable	Poor	High**
Method convenient?	Yes	Moderately	No	Yes	No
DNA storage	Long-term	Long-term	Short-term	No	Long-term
Performance in downstream applications	Excellent	Moderate	Poor	Very poor	Excellent

* It is recommended that plants are grown in the dark for 2 days before isolating DNA.
† Protocol may need to be optimized for some species.
‡ Using the DNeasy Plant Mini Kit.
§ Does not include the recommended 2 day dark treatment.
** Depends on handling.

Genetic Markers, Map Construction, and Their Application in Plant Breeding

Jack E. Staub[1] and Felix C. Serquen[2]

Vegetable Crops Research, U. S. Department of Agriculture, Agricultural Research Service, Department of Horticulture, University of Wisconsin–Madison, WI 53706

Manju Gupta[3]

Mycogen Plant Sciences, Madison Laboratories, 5649 East Buckeye Road, Madison, WI 53716

The genetic improvement of a species through artificial selection depends on the ability to capitalize on genetic effects that can be distinguished from environmental effects. Phenotypic selection based on traits that are conditioned by additive allelic effects can produce dramatic, economically important changes in breeding populations. Genetic markers—heritable entities that are associated with economically important traits—can be used by plant breeders as selection tools (Beckman and Soller, 1983; Darvasi and Soller, 1994). Marker-assisted selection (MAS) provides a potential for increasing selection efficiency by allowing for earlier selection and reducing plant population size used during selection. Nevertheless, the phenotypic variation that marker loci define is often nonadditive, and is a function of genetic linkage, pleiotropy, and environment (Lark et al., 1995). Thus, the efficiency of application of marker loci as predictors of phenotypic variation depends on many factors, and predictions of response to selection (R) or genetic gain (Δ G) are often difficult.

The predictive value of genetic markers used in MAS depends on their inherent repeatability (Weeden et al., 1992), map position, and linkage with economically important traits (quantitative or qualitative). The presence of a tight linkage (<10 cM [centimorgan]) between qualitative trait(s) and a genetic marker(s) may be useful in MAS to increase gain from selection (Kennard et al., 1994; Paran et al., 1991; Timmerman et al., 1994). Likewise, selection for multiple loci or quantitative trait loci (QTL) using genetic markers can be effective if a significant association is found between a quantitative trait and markers (Edwards and Page, 1994; Edwards et al., 1987; Lande and Thompson, 1990).

Often the biotechnological information presented in research reports is not tied directly to classical genetic methodologies and the sophisticated technology presented results in a bewildering array of new terms. For scientists who have a peripheral interest in genome mapping, but would like to understand the potential role of MAS in plant improvement, the wealth of information currently being produced in this area can lead to considerable confusion. The purpose of this paper is to describe available marker types and examine factors critical for their use in map construction and MAS. This review clarifies how genetic markers are used in map construction and defines the potential use of genetic maps for MAS.

Received for publication 14 July 1995. Accepted for publication 24 Nov. 1995. Mention of a trade name, proprietary product, or specific equipment does not constitute a guarantee or warranty by the USDA and does not imply its approval to the exclusion of other products that may be suitable. The cost of publishing this paper was defrayed in part by the payment of page charges. Under postal regulations, this paper therefore must be hereby marked *advertisement* solely to indicate this fact.

[1]Research Horticulturist and Professor.

[2]Research Associate.

[3]Research Scientist, Applied Genetics.

MARKER TYPES

Morphological. Morphological traits controlled by a single locus can be used as genetic markers if their expression is reproducible over a range of environments. Although codominant morphological markers have been useful as predictors of genetic response to selection, they can be influenced by environmental and genetic factors (e.g., epistasis). For instance, the expression of the determinate (*de*) character in cucumber (*Cucumis sativus* L.) may vary, depending on growing environment and modifying genes (Staub and Crubaugh, 1995). Thus, a description of such a trait has significance only when accompanied by properly documented pedigree information and environmental conditions. The fact that such factors may modify a gene's expression of phenotype may limit its usefulness as a genetic marker. A further drawback of morphological markers is that they may present an altered phenotype that interferes with grower needs.

Isozymes. Isozymes are differently charged protein molecules that can be separated using electrophoretic procedures (usually starch gel) (Markert and Moller, 1959). Since enzymes catalyze specific biochemical reactions, it is possible to visualize the location of a particular enzyme on a gel by supplying the appropriate substrate and cofactors, and involving the product of the enzymatic reaction in a color-producing reaction. The colored product becomes deposited on the gel, forming a visible band where a particular enzyme has been electrophoretically localized. Bands visualized from specific enzymes represent protein products, have a genetic basis, and can provide genetic information as codominant markers. However, the paucity of isozyme loci and the fact that they are subject to post-translational modifications often restricts their utility (Staub et al., 1982).

RFLPs. Restriction fragment length polymorphisms (RFLPs) are detected by the use of restriction enzymes that cut genomic DNA molecules at specific nucleotide sequences (restriction sites), thereby yielding variable-size DNA fragments (Fig. 1). Identification of genomic DNA fragments is made by Southern blotting, a procedure whereby DNA fragments, separated by electrophoresis, are transferred to nitrocellulose or nylon filter (Southern, 1975). Filter-immobilized DNA is allowed to hybridize to radioactively labeled probe DNA. Probes are usually small [500 to 3000 base pairs (bp)], cloned DNA segments (e.g., genomic or cDNA). The filter is placed against photographic film, where radioactive disintegrations from the probe result in visible bands. Such bands are visualizations of RFLPs, which are codominant markers.

The polymerase chain reaction (PCR) has been used to develop several DNA marker systems (Fig. 1). Three strategies primarily have been employed in the development of PCR-based marker systems. These include: 1) markers that are amplified using single primers in PCR, where marker system diversity results from variation in the length and/or sequence of primers, and where anchor nucleotides are present at 5′ or 3′ termini of primers (e.g., RAPDs, SPARs, DAFs, AP-PCR, SSR-anchored PCR; see below); 2) markers that are selectively amplified with two primers in PCR such that their selectivity comes from the presence of two to four random bases at the 3′ ends of primers that anneal to the target DNA during the PCR (e.g., AFLP, below); and 3) markers amplified using two primers in PCR, which commonly requires cloning and/or sequencing for the construction of specific primers. In this case, variations in marker technology result from differences in the target DNA sequence present between two primers (e.g., AMP-FLPs, STRs, and SSRs).

RAPDs. Of three similar, single-primer, PCR-based technologies, random amplified polymorphic DNA (RAPD; Williams et al., 1990), DNA amplification fingerprinting (DAF; Caetano-Anolles et al., 1991), and arbitrary-primed PCR, (AP-PCR; Owen and Uyeda, 1991; Welsh and McClelland, 1990), RAPDs have been used most widely for map construction and linkage analysis (Reiter et al., 1992) (Fig. 1). RAPD markers are generated by PCR amplification of random genomic DNA segments with single primers [usually 10 nucleotides (nt) long] of arbitrary sequence (Williams et al., 1990). The primer/target complexes are used as substrates for

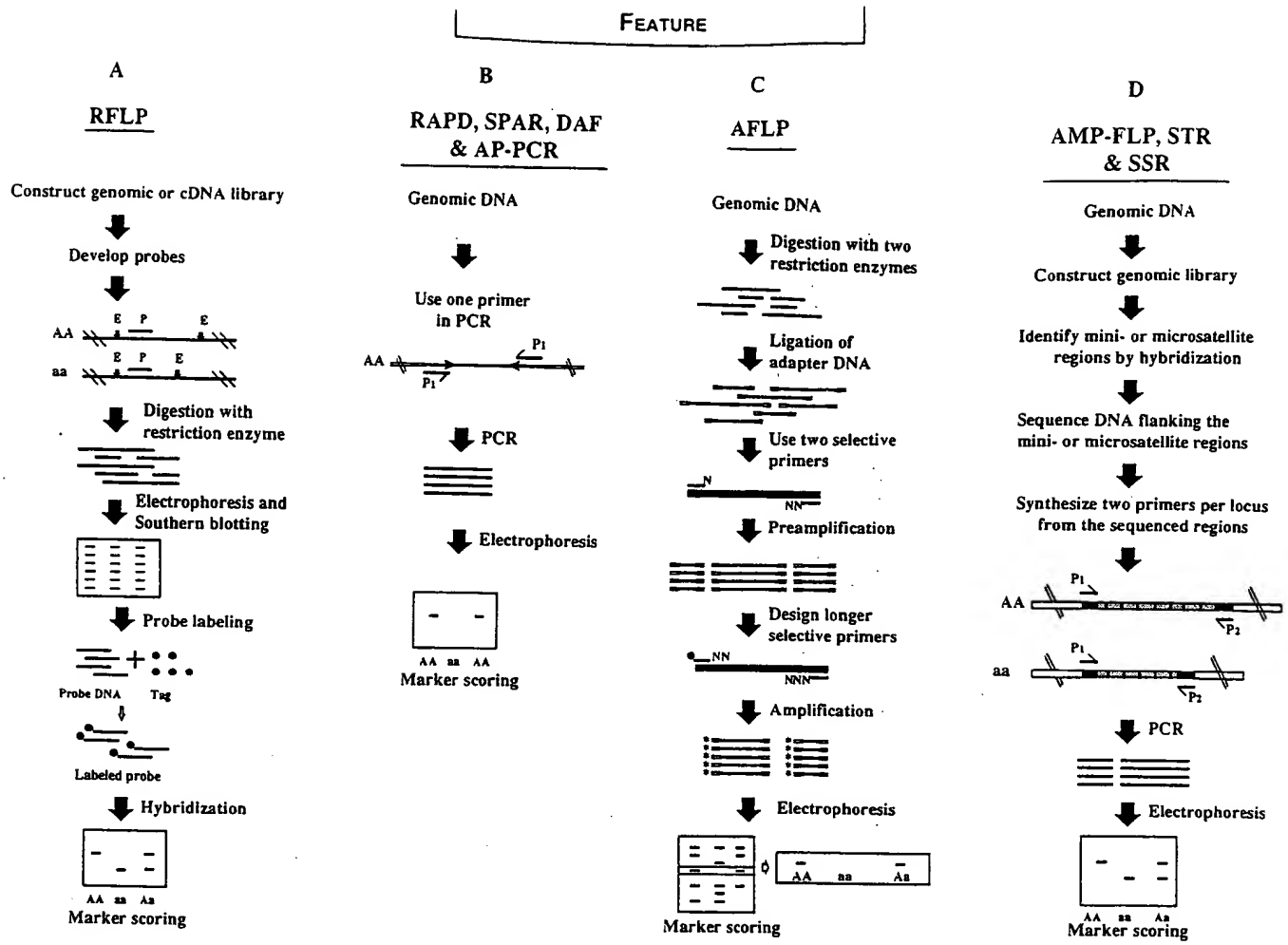


Fig. 1. Comparisons of various molecular marker technologies. In contrast to technologies presented in panels B and C, those in panels A and D require an expensive and time-consuming set-up phase for probe and primer development. Based on their availability, probes and primers can also be obtained from institutions or universities. DNA markers in panels B and C can be generated within a short time and require minor development. (A) E = restriction enzyme, P = probe, tag = either radioactive or nonradioactive for detection of hybridized probes. (B) P_1 = single primer of polymerase chain reaction (PCR). Although most PCR-based technologies amplify few to multiple bands, examples of only one band are given to illustrate scoring. (C) Adapter DNA = short double-stranded DNA molecules, 18 to 20 base pairs in length, representing a mixture of two types of molecules. Each type is compatible with one E generated DNA ends. Selective primers = primers for preamplification designed to contain an adapter DNA sequence plus one to two random bases at the 3′ end for reading into the genomic fragments. Primers for amplification have the preamplification primer sequence plus one to two additional bases at the 3′ end. A tag is attached at the 5′ end of one of the amplification primers for detecting amplified molecules (multiple bands are also possible). Differences in scoring exist between bands in this panel and those of panel B (also see Fig. 2). (D) P1 and P2 = two PCR primers.

DNA polymerase to copy the genomic sequences 3′ to the primers. Iteration of this process yields a discrete set of amplified DNA products that represent target sequences flanked by opposite-oriented primer annealing sites. Amplification products can be separated by electrophoresis on agarose or polyacrylamide gels and visualized by staining with ethidium bromide or silver. RAPDs are usually dominant markers with polymorphisms between individuals defined as the presence or absence of a particular RAPD band (Fig. 2). The further development of RAPD methodology has produced other PRC-based markers (e.g., SCAR and ASAP markers).

SCARs. Utility of a desired RAPD marker can be increased by sequencing its termini and designing longer primers (e.g., 24 nt) for specific amplification of markers (Paran and Michelmore, 1993). Such sequenced characterized amplified regions (SCARs) are similar to sequence-tagged sites (STS) (Olson et al., 1989) in construction and application. DNA sequence differences are manifest by the presence or absence of a single unique band. SCARs are more reproducible than RAPDs and can be developed into plus/minus arrays where electrophoresis is not needed. Although SCARs are usually dominant markers, some SCARs can theoretically be converted to codominant markers by digestion with 4-bp restriction endonucleases, and identification of polymorphisms by either denaturing gradient gel electrophoresis (DGGE) or single-strand conformational polymorphism (SSCP) techniques (Rafalski and Tingey, 1993).

ASAPs. A recent modification of PCR technology involves the alkaline extraction of DNA with subsequent amplification of the DNA template in microtiter plates using allele-specific associated primers (ASAPs) that generate only a single DNA fragment at stringent annealing temperatures (similar to SCARs) (Gu et al., 1995). The DNA fragment is present in only those individuals possessing the appropriate allele and thus eliminates the need to separate amplified DNA fragments by electrophoresis (i.e., presence/absence polymorphism). This method involves ethidium bromide binding to the DNA double helix, which dramatically enhances its fluorescence but does not bind to free nucleotides in the PCR mixture. This approach was developed to decrease time for DNA extraction and increase the reliability of the PCR reaction for large-scale screening.

SPARs. The single primer amplification reaction (SPAR) is a DNA marker system that can produce multiple markers per assay (Fig. 1). The system uses primers based on microsatellites or simple sequence repeats (SSRs) and amplifies inter-SSR DNA sequences (Gupta et al., 1994). Of the di-, tri-, tetra-, and pentanucleotide SSRs, the tetranucleotide repeats are most effective in producing polymorphic multiband patterns. The level of polymorphism is related to the genomic diversity within a given species. Most DNA markers map to scattered genomic locations. Although most SSR-SPARs are dominant markers, codominant markers can also be detected (Fig. 2). Given that an unlimited number of primers can be synthesized from the tetranucleotide repeats [$(4)^4 = 256$], and from the combination of di-, tri-, and

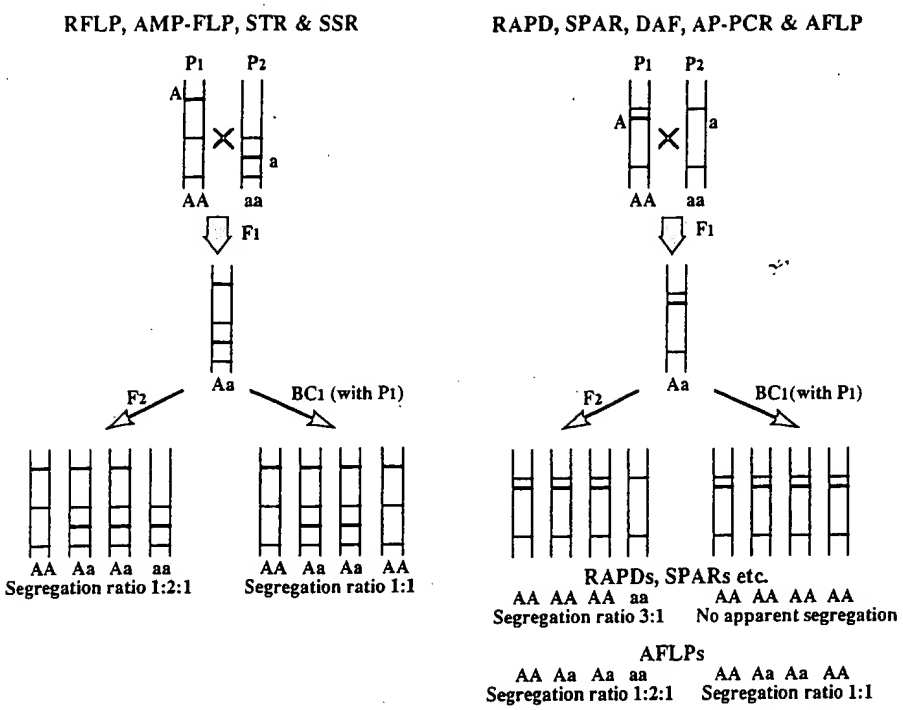


Fig. 2. Schematic of inheritance patterns of DNA markers in F_2 and BC_1 progenies. Common bands shown in the patterns of the parents and progeny may not be seen with other probes and primers. RFLPs, AMP-FLPs, STRs, and SSRs are codominant markers and thus heterozygousity in F_2 and BC_1 is easily detected. In contrast, RAPDs, SPARs, DAPs, AP-PCR, and AFLPs are dominant markers and detection of heterozygosity is, with rare exception, not possible. Zygosity determination is also possible through the quantification of DNA bands (e.g., AFLPs).

tetranucleotide SSRs, or compound SSRs, the SSR-SPAR marker system may have broad application across a range of plant species.

SSR-anchored PCR. This system employs single primers of dinucleotide simple sequence repeats (SSRs; see below); especially (CA)n repeats for amplification of markers. The primer is either anchored at 3′ or 5′ termini with two to four nucleotides (Zietkiewicz et al., 1994). Multiple bands containing inter-SSR regions are amplified and then are fractionated on polyacrylamide gels for pattern visualization. These amplified bands are mostly dominant markers and can be used in a wide range of plant species.

AFLPs. Production of amplified fragment length polymorphisms (AFLPs) is based on selective amplification of restriction enzyme-digested DNA fragments (Zabeau and Vos, 1993) (Fig. 1). Multiple bands are generated in each amplification reaction that contains DNA markers of random origin. Analysis of DNA on denaturing polyacrylamide gels typically results in the production of 50 to 100 bands per individual sample. AFLPs are quantitative in that heterozygous and homozygous genotypes can be differentiated by the intensity of the amplified bands. The ability of this technology to generate many markers with minimum primer testing, and the system's high resolution (i.e., band clarity and relatively low lane background) are features that make AFLPs attractive as genetic markers (primarily dominant; Fig. 2). Because of its expense, automation may be required to realize this technology's full potential during MAS.

AMP-FLPs, STRs, and SSRs. Mini- and microsatellite DNA sequences are an excellent source of polymorphisms in eukaryotic genomes, and are well suited for genotyping and map construction. Marker systems based on such sequences include amplified fragment length polymorphisms (AMP-FLPs; minisatillites in vertebrates; Fregeau and Fourney, 1993), short tandem repeats (STRs; microsatellites in verebrates; Fregeau and Fourney, 1993), and simple sequence repeats (SSRs; microsatellites in plants; Rafalski and Tingey, 1993) (Fig. 1). Mini- and microsatellites are comprised of tandem arrays of 15- to 70-bp and 2- to 5-bp monomeric repeat units, respectively. Polymorphisms appear because of variation in the number of tandem repeats in a given repeat motif. Most STRs and SSRs are dinucleotide repeat-based [(AC)n, (AG)n, and (AT)n] microsatellite markers (Rafalski and Tingey, 1993). Such polymorphisms are amplified by designing primers from the sequenced regions flanking the repeat motifs (Fig. 1). Similar to (CA)n repeats in humans, (AT)n dinucleotide microsatellite repeats are relatively abundant and highly polymorphic in plants (Akkaya et al., 1992; Morgante et al., 1994). This group of markers is codominant in its expression (Fig. 2).

CAPs. Cleaved amplified polymorphic sequences (CAPs) are a form of genetic variation in the length of DNA fragments generated by the restriction digestion of PCR products (Koniecyzn and Ausubel, 1993; Jarvis et al., 1994). The source of the sequence information for the primers can come from a genebank, genomic or cDNA clones, or cloned RAPD bands. This marker class is codominant in its behavior.

MARKER SYSTEM SELECTION

Selection of a DNA marker system for plant breeding depends on project objectives, population structure, the genomic diversity of the species under investigation, marker system availability, time required for analysis, and the cost per unit information (Table 1). Clearly, each marker system has advantages and disadvantages, and therefore it is critical to evaluate each marker system for its potential utility before use. For example, intraspecific maps can be constructed with a common set of RFLP probes; however, each species initially requires the construction of a map. The range of polymorphism in species also plays a role in marker selection (e.g., in self-pollinated species RAPDs are more useful for detecting polymorphisms within a gene pool than RFLPs). Moreover, the use of a marker system in one species does not necessarily indicate its efficacy in another species.

Marker systems also differ in their utility across populations, species, and genera, and their efficacy in the detection of polymorphism. For instance, RFLPs mapped in one population can be used as probes for characterizing other populations within the same species. In contrast, SSRs can be as informative as RFLPs, but polymorphic primers identified in one species are generally not useful in another species. Likewise, maps using RAPDs, SPARs, and AFLPs can be constructed in a relatively short period; however, such markers are not useable across populations, because each marker is primarily defined by its length (i.e., sequence information may be limited). Moreover, the same size band amplified across populations/species does not necessarily mean that bands possess the same sequence, unless proven by hybridization studies (Thormann et al., 1994). In contrast to RFLPs, these marker systems possess all the advantages of PCR-based systems (i.e., small sample requirement, high throughput, and early selection) (Table 1). These advantages can be nullified if polymorphism within a species is low. Where the level of polymorphism is low, STRs and SSRs are currently the marker systems of choice. However, the cost and time required to develop these marker systems can be considerable (Table 1).

Costs per unit information (data point) depend on the time required for DNA extraction (sampling), the amount of DNA needed for analysis, whether cloning and sequencing is necessary, the amount of potentially useful genetic information acquired, the type of genetic information needed, whether the allelic variation can be ascribed to banding patterns (dominant vs. codominant), whether the electrophoretic system can be automated, the potential utility of genetic maps, and the proprietary status of the technique (Table 1). Codominant markers, such as RFLPs, are useful for MAS and evolutionary studies, but their use can be time consuming, relatively expensive, and may require considerable technical exper-

Table 1. Comparisons among several molecular marker systems for various technical attributes and proprietary rights status.

Critical variables	Molecular marker systems[z]					
	RFLP	AFLP	SPAR	RAPD	DAF	SSR/STR/AMP-FLP
Tissue sampling (weeks)[y]	4	2	2	2	2	2
DNA needed/100 markers (mg)	100	0.5	2.5	2.5	2.5	10
Cloning and sequencing	Yes	No	No	No	No	Yes
Information content per run[x]	0–3	0–30	0–20	0–4	0–10	0–2
Marker type[w]	C	D	D	D	D	C
Zygosity detection[v]	Yes	Yes	No	No	No	Yes
Automation[u]	+	+++	++	++	++	+++
Utility of genetic maps[t]	SS	CS	CS	CS	CS	SS
Proprietary rights status[s]	NA	LC	NA	LC	LC	NA

[z]RFLP = restriction fragment length polymorphism, AFLP = amplified fragment length polymorphism, SPAR = single primer amplification reaction, RAPD = random amplied polymorphic DNA, DAF = DNA amplification fingerprinting, and SSR/AMP-FLP = simple sequence repeats/amplified fragment length polymorphism.

[y]The sampling time after sowing is shown for corn (*Zea mays*) (relative time applies to all crop species).

[x]Markers obtained per hybridization or PCR reaction. For SSR/AMP-FLP, number of markers per run does not reflect multiplexing.

[w]D and C equal dominant and codominant markers, respectively.

[v]Heterozygous alleles can be distinguished from the homozygous alleles.

[u]On the scale of 1 to 3, + = the least and +++ shows the most potential for automation. Automation refers to mechanizing steps involving processing of DNA and detection, identification and scoring of markers.

[t]Refers to the relative utility of maps constructed with a given marker system either within that species (SS = species specific) or to a specific cross or population (CS = cross specific).

[s]NA = not applicable, LC = license required to practice the technology.

tise. Often the high cost of developing an RFLP marker system for a new species or inefficiency of MAS, due to the large quantity of DNA required and slow screening process, results in a decision to use a PCR-based system. Nevertheless, RFLPs command an advantage in systematic and evolutionary studies because, in contrast to most PCR-based technologies, which detect variation in 20–40 bases (combined length of primers), data aquisition is based on the homology among large fragments of DNA (length of probes). The use of RFLPs for such studies is enhanced if polymorphisms are abundant.

MAP CONSTRUCTION

The development of molecular marker technology and consequent identification of many marker loci has caused renewed interest in genetic mapping. Genetic map construction requires that the researcher: 1) select the most appropriate mapping population(s); 2) calculate pairwise recombination frequencies using these population(s); 3) establish linkage groups and estimate map distances; and 4) determine map order. Since large mapping populations are often characterized by different marker systems, map construction has become computerized. Computer packages such as Linkage 1 (Suiter et al., 1983), GMendel (Echt et al., 1992), Mapmaker (Lander and Botstein, 1986; Lander et al., 1987), MapManager (Manly and Elliot, 1991), and JoinMap (Stam, 1993) have been developed to aid in the analysis of genetic data for map construction. These programs use data obtained from segregating populations to estimate recombination frequencies that are then used to determine the linear arrangement of genetic markers by minimizing recombination events.

Mapping populations

Selection of mapping populations is critical to successful map construction. Since a map's economic significance will depend upon marker-trait associations, as many qualitatively inherited morphological traits as possible should be included in genetic stocks chosen as parents for map construction (Table 2). Consideration must be given to the source of parents (adapted vs. exotic) used in the mapping population. Chromosome pairing and recombination rates can be severely disturbed (suppressed) in wide crosses (adapted x exotic) and generally yield greatly reduced linkage distances (Albini and Jones, 1987; Zamir and Tadmor, 1986). Wide crosses will usually provide segregating populations with a relatively large array of polymorphisms when compared to progeny segregating in a narrow cross (adapted x adapted). To have significant value in plant improvement programs, a map made from a wide cross must be colinear (i.e., order of loci similar) with maps constructed using adapted parents.

The choice of an appropriate mapping population depends on the type of marker systems employed (Tanksley et al., 1988). Maximum genetic information is obtained from a completely classified F_2 population using a codominant marker system (Mather, 1938). Information from a dominant marker system can be equivalent to a completely classified F_2 population if progeny tests (i.e., F_3 or F_2BC) are used to identify heterozygous F_2 individuals. This procedure is often prohibitive because of the cost and time involved in progeny testing.

Dominant markers supply as much information as codominant markers in recombinant inbred lines (RI) (i.e., an array of genetically related lines; usually $>F_8$), doubled haploids, or backcross populations in coupling phase (Burr et al., 1988). Information obtained from dominant markers can be maximized by using RI or doubled haploids because all loci are homozygous, or nearly so. Under conditions of tight linkage (i.e., about <10% recombination), dominant and codominant markers evaluated in RI populations provide more information per individual than either marker type in backcross populations (Reiter et al., 1992). However, as the distance between markers becomes larger (i.e., loci become more independent), the information obtained per unit individual in RI populations decreases dramatically when compared to codominant markers.

Backcross populations can be useful for mapping dominant markers if all loci in the recurrent parent are homozygous and the donor and recurrent parent have contrasting polymorphic marker alleles (Reiter et al., 1992). Information obtained from backcross populations using either codominant or dominant markers is less than that obtained from F_2 populations because one, rather than two, recombinant gametes are sampled per plant. Backcross populations, however, are more informative (at low marker saturation) when compared to RIs as the distance between linked loci increases in RI populations (i.e., about >15% recombination). Increased recombination can be beneficial for resolution of tight linkages, but may be undesirable in the construction of maps with low marker saturation.

Progeny testing of F_2 individuals is often used in map construction where phenotypes do not consistently reflect genotype (e.g., disease reaction and many useful traits) or where trait expression is controlled by QTLs. Segregation data from progeny test populations (e.g., F_3 or F_2BC) can be used in map construction. MAS can then be applied to cross progeny based on marker-trait map associations, especially in early generations (F_2, F_3), where linkage groups have not been completely disassociated by recombination events (i.e., maximum disequilibrium).

Recently, a method has been developed for the rapid identification of linkage using bulked segregant analysis (BSA; Michelmore et al., 1991). In BSA, two bulked DNA samples are drawn from a segregating population originating from a single cross. These bulks contain individuals that are identical for a particular trait (e.g., resistant or susceptible to a particu-

lar disease) or genomic region but arbitrary at all unlinked regions (i.e., heterozygous). The bulks are screened for DNA polymorphisms and these differences compared against a randomized genetic background of unlinked loci. Thus, differences between the two bulks indicate markers (e.g., bands) that are linked to a particular trait. BSA overcomes several problems that are associated with the use of nearly isogenic lines (NILs), which require many backcrosses to develop. Where only a portion of the polymorphic loci are expected to map to a selected region using NILs (e.g., BC_5 only 50%), regions unlinked to the target region will not differ between the bulked samples of many individuals in BSA. Moreover, all loci detected during BSA will segregate and can be mapped, thus eliminating the linkage drag problems (i.e., genes incorporated into lines by backcrossing that are flanked by DNA segments introduced from the donor parent) associated with NILs (Young and Tanksley, 1989).

Calculation of recombination fraction

Crossover events can be described as the percentage of recombination in offspring. Only half the meiotic products will be crossover types (recombinants) when one chiasma forms between two loci. Multiple crossovers can also be detected through observation of progeny phenotypes. Single crossover events are not independent and the number of double crossover events is usually smaller than predicted. This positive "interference" varies, depending on organism, crossover location, environmental factors, and numerous other factors. Therefore, accurate estimates of double crossing over can only be obtained when interference is considered. Interference is measured as a coefficient of coincidence (CC), which is an expression of the ratio of observed double crossovers to those predicted (expected) by a map. The expected double crossover frequency is predictable if two crossovers are independent events or if interference can be measured.

The proportion of mean number of recombination events defines the map distance between two loci. The relationship between map distance and recombination value is characterized by a genetic mapping function (*mf*). An *mf* is a formula expressing quantitative relationships between distances in a linkage map using crossover frequency. There are several types of mapping functions that can be applied, depending on the assumed degree of crossover interference that best represents the mapping population. The most common mapping functions were developed by Haldane (1919) and Kosambi (1944). While Haldane's *mf* assumes absence of interference, Kosambi's assumes positive interference (i.e., fewer double recombinants when compared to no interference).

The frequency of recombinant gametes produced can be used as an index of the distance between two loci on a chromosome [1 map unit = about 1 cM]. Map distance, however, is not completely additive. Additivity is based on the assumption that the average number of crossovers per chromatid occurring between two loci is directly proportional to the distance between the two loci. The frequency of recombination (percentage) and map distance is, however, not directly proportional. Estimates of the frequency of crossing over will be most reliable when genes are relatively closely linked (1 to 10 map units). Recombination percentage is only equivalent to map distance within the range of the minimum distance for crossing over (lack of additivity) because double crossovers occur at significant frequency. A nonlinear relationship occurs

Table 2. Total map length and mean distance between genetic markers in various plant species.

Common name	Botanical name	Marker			Map		Reference
		Pop.[z]	Type[y]	No.	Length (cM)	Mean distance (cM)	
Arabidopsis	*A. thaliana* L.	RI	RFLP	320	630	2.0	Reiter et al., 1992
Banana	*Musa acuminata* Colla	F_2	RFLP, isozyme, RAPD	90	606	10.0	Faure et al., 1993
Barley	*Hordeum vulgare* L.	DH	RFLP, isozyme, RAPD, morphological, SAP, disease	295	1250	4.2	Kleinhofs et al., 1993
Bean	*Phaseolus vulgaris* L.	BC	RFLP, isozyme, seed protein, flower color	244	1200	≈5.0	Vallejos et al., 1992
		F_2	RFLP, RAPD, isozyme	152	827	6.5	Nodari et al., 1993a
Brassica	*B. napus* L.	F_2	RFLP, isozyme, RAPD, morphology, disease	120	1413	14.0	Landry et al., 1991
	B. napus L.	F_1-DH	RFLP	132	1016	7.7	Ferreira et al., 1994
	B. rapa L. (syn. *campestris*)	F_2	RFLP	280	1850	6.9	Song et al., 1991
	B. rapa	F_3	RFLP, seed color, seed erucic acid, pubescence	139	1785	13.5	Teutonico & Osborn, 1994
Citrus	*C. grandis* L. x *C. paradisi* Maef.	F_1	RFLP	46	1700	20.0	Jarrell et al., 1992
Cotton	*Gossypium hirsutum* L. x *G. barbadense* L.	F_2	RFLP	705	4675	7.1	Reinisch et al., 1994
Cucumber	*Cucumis sativus* L.	F_2	RFLP, RAPD, isozyme, morphology, disease	58	766	8.0	Kennard et al., 1994
	C. sativus x *C. hardwickii* (R.) Alef.	F_2	RFLP, morphology	70	480	8.1	
Cuphea	*Cuphea lanceolata* Ait.	F_2	RFLP	37	288	7.8	Webb et al., 1992
Maize	*Zea mays* L.	RI	RFLP	334	1460	≈5.0	Burr et al., 1988
Potato	*Solanum tuberosum* L. x *S. berthaultii* Corr.	BC	RFLP, isozyme	977	684	0.7	Tanksley et al., 1992
Rice	*Oryza sativa* L. x *O. longistaminata* A. Chev. & Roehr	BC	RFLP	726	1491	2.0	Causse et al., 1994
Rye	*Secale cereale* L.	IBL	RFLP, isozyme, RAPD, morphology, physiology	60	350	6.0	Philipp et al., 1994
Sorghum	*S. bicolor* L.	F_2	RFLP	98	949	10.0	Whitkus et al., 1992
		F_2	RFLP	190	1789	9.4	Xu et al., 1994
Soybean	*Gylcine max* L.	F_2	RFLP	252	2147	8.5	Diers et al., 1992
Sugar beet	*Beta vulgaris* L.	F_2	RFLP	115	789	6.9	Pillen et al., 1992
Tomato	*Lycopersicon esculentum* Miller x *L. pennelli*	F_2	RFLP, isozyme	1030	1276	1.2	Tanksley et al., 1992

[z]RI = recombinant inbred, DH = doubled haploid, and IBL = inbred line.
[y]RFLP = restriction fragment length polymorphism, RAPD = random amplified polymorphic DNA, and SAP = specific amplicon polymorphism.

when estimates are made outside of this distance. Therefore, the actual map distance between two genes will tend to be underestimated by the recombination fraction (r) (e.g., $r = 0.10 = 10$ cMs, $r = 0.30 = 45.8$ cMs, $r = 0.35 = 60.2$ cMs), such that at large distances (~40–50 map units) the two genes will be strictly independent of each other (Kosambi, 1944).

Linkage phase

Genes are linked when they are on the same chromosome. There are two possible arrangements of two genes on a pair of chromosomes—coupling and repulsion. Coupling signifies that the two recessive alleles are carried in one chromosome and the two dominant alleles in the other (i.e., *AB/ab*); repulsion describes the alternate arrangement (i.e., *Ab/aB*). This relationship is particularly important when dealing with dominant markers during map construction. Two linked markers scorable as dominant alleles [e.g., *AA* or band presence (+)] can only be recognized in coupling phase linkage. This is because the heterozygote class cannot be distinguished from the homozygote dominant class (i.e., band presence = *AA* or *Aa*, band absence = *aa*). In contrast, codominant markers allow for the expression of both pairs of alleles (i.e., phenotypically as *AA*, *Aa*, *aa*). Linkage phase has proven important in selection when dominant markers are used (Haley et al., 1994). A greater proportion of bean (*Phaseolus vulgaris* L.) genotypes homozygous resistant to bean common mosaic virus (BCMV), and a lower proportion of segregating and homozygous susceptible genotypes were recovered when selection was imposed against a repulsion-phase RAPD marker than when selection was made for the coupling-phase RAPD marker. This observation is of practical significance where resistance is conditioned by recessive genes since it requires breeders to select against the heterozygous susceptible individuals (Kelly, 1995). In the case of selection for the recessive BCMV resistance gene (*bc-3*) in bean, Kelly stated: "selection of individuals based on the phenotype of combined coupling and repulsion-phase RAPD markers was equivalent to selection based on a codominant marker (RFLP) and was identical to selection based on the repulsion-phase marker alone."

Establishment of linkage groups

If linkage is indicated by Chi-square analysis of progeny segregation, then the potential for linkage between loci can be mathematically tested. There are several mathematical methods available for investigating potential linkage relationships (Crow, 1990). Among these are the maximum likelihood and least squares/regression methods. Currently, these are the methods of choice for linkage estimation because they result in estimates that have the smallest standard error (Mather, 1938; Nordheim et al., 1984). They are especially useful where multiple loci (QTL) are involved (Shute, 1988). While least squares estimation attempts to minimize deviations from a mathematical model (regression), maximum likelihood involves comparisons among two or more plausible hypotheses (e.g., linkage vs. no linkage). The maximum likelihood method is particularly useful in evaluating genetic phenomena and will be used in this discussion as an illustration of linkage analysis (Chakravarti et al., 1991).

Maximum likelihood and likelihood odds ratio (LOD) value

Mather (1938) developed the maximum likelihood approach for linkage analysis. It is used by various computer-based linkage programs (e.g., Mapmaker; Lander et al., 1987) to determine the probability of linkage between a given marker and a known marker. Maximum likelihood is a statistical procedure designed to choose values for variables that maximize a defined function, which is done by integrating the function and solving for 0 (minimizing the integrated function) or by iteration.

Linkage estimation using the method of maximum likelihood is based on the binomial expansion, which is a special case of the polynomial $(m_1 + m_2 + m_3 + \ldots)^n$. Maximum likelihood, as applied to linkage estimation, attempts to select a linkage estimator (r value) that minimizes an expectation function in a binomial expression. The benefit of maximum likelihood in the calculation and estimation of r is that functions can be designed that include ambiguous classes. An example of an ambiguous class is the double heterozygote (*AaBb*) of an F_2 family that contains recombinant and non-recombinant types.

Recombination value is used in the maximization expression to determine the likelihood (L) of association of linkage between a set of variables (i.e., genetic loci). The value of r that maximizes the likelihood of the observed outcome is determined. Solutions are limited to the range of 0 to 0.5.

After maximization, the question is raised as to whether the value of r, say x, is significant, given the upper limit of no linkage (i.e., 0.5)—that is, whether the probability that two loci are linked with a given r value over the probability that the two loci are not linked. An understanding of the precision of r is necessary to assess the utility of the value obtained by likelihood maximization. Historically, this has been done in two ways. Allard (1956) constructed a series of tables and formulae to calculate recombination values and associated standard errors using a maximum likelihood approach. This approach had already been widely used by plant geneticists (Fisher, 1946; Kramer and Burnham, 1947). Researchers in human genetics gravitated toward the use of the LOD defined by Haldane and Smith (1947). This approach has been used by some plant researchers because the LOD calculations needed for analyzing large populations and using many markers has been simplified by the use of some computer-based linkage estimation programs (e.g., Mapmaker; Lander and Botstein, 1986, Lander et al., 1987).

The odds ratio of a maximization event is given as: L (x)/L (0.5). This form, however, is inconvenient in most instances and the log of the odds ratio {i.e., LOD = log [L (x)/L (0.5)]} is used (Risch, 1992). In many analyses, a significance level of LOD > 3.0 is appropriate as an acceptance level of linkage between two loci. This value is equivalent to saying that the alternative hypothesis (linkage) has to be greater than 1000 times more probable than the null (no linkage) hypothesis. If this analysis is repeated over 100 marker loci, a significant level of LOD > 3 for each locus is comparable to an experiment-wise (genome-wise) type I error rate of alpha (α) = 0.01. LOD decreases with increasing r values and increases with increasing sample size.

Tests of linkage for qualitatively inherited traits vary in scope and operation. The researcher must determine a threshold LOD value below which linkage is not considered significant (Churchill and Doerge, 1994). As the LOD threshold is raised, fewer markers are assigned to linkage groups (i.e., independent loci), and more and smaller linkage groups are identified. Comparison of maps created from an array of LOD values often allows the researcher to determine the stability of putative linkage groupings. It is clear that any map only approximates reality and that map distances between markers will change as new information (i.e., more markers) becomes available.

Gene order determination and map merging

Because additivity of map distance is accepted (assuming no double crossovers occur) for narrow intervals (1–10 map units), tightly linked genes can be placed in relative order. Genes that are loosely linked (>20 map units) can be placed on a map but their location is much more tentative. The map distances calculated based on crossover percentages (i.e., genetic map) often bear no direct relationship with the actual physical distances between linked genes (i.e., physical map) (Stansfield, 1969; Swanson et al., 1990). The linear order in the physical and genetic maps, however, should theoretically be identical.

Three linked genes may be in any one of three orders, depending on which gene is in the middle of the linkage group. Traditionally, gene orders have been determined from either two- or three-point testcross data. When multiple crossovers occur with much greater than random frequency (i.e., localized negative interference), gene order of closely linked sites can be ascertained using three-factor reciprocal crosses.

Genetic maps in several crop species have been constructed using various marker systems, types of populations, and, often, generations. Although selected data for several crops are presented for comparison (Table 2), detailed and updated information on these and other species resides in genome databases housed in the U.S. National Agricultural Library, Beltsville, Md. Maps in many species are moderately saturated and incorporate isozymes, RFLPs, and RAPDs. There are at least three maps for potato (*Solanum tuberosum* L.) and rice (*Oryza sativa* L.), and two for bean

and pepper (*Capsicum annuum* L.) that have been constructed using various parents analyzed in diverse generations. Likewise, multiple maps are being developed for other plant species [e.g., *Arabidopsis thalania* (L.) Heynh., corn (*Zea mays* L.), and tomatoes (*Lycopersicon esculentum* Mill.)].

If two or more genetic maps possess a minimal number of common markers they can be merged to create a more informative map (Hauge et al., 1993). However, the type of information (e.g., F_2 vs. BC) and precision of estimates of recombination frequencies (family size) often vary greatly between populations and data sets. Therefore, any procedure that attempts to merge mapping information must "weigh" these types of information to create the "optimal," "most likely" map with the least amount of "internal tension."

A computer program, JoinMap, has recently been developed that considers the estimates of recombination frequency between a given pair of markers of different origin (data sets/mapping populations), calculates and applies the appropriate weighting, and then generates a single recombination value (Stam, 1993). After assigning weights to all available pairwise combinations, JoinMap institutes a numerical search for the best-fitting linear arrangement of the marker loci. JoinMap calculates a goodness-of-fit criterion corresponding to the two hypothesized levels of interference (positive and negative), allowing for an examination of each synthesized map.

Identification of QTLs

In contrast to classical linkage detection for single gene traits, different strategies have been suggested for the identification (i.e., detection and localization) of single QTLs (Edwards et al., 1987; Jiang and Zeng, 1995; Lander and Botstein, 1989). Such strategies attempt to identify major levels of the total genetic variance that contribute to a trait's variation. They differ in approach in the number of markers that they evaluate during linkage estimation. Tests for QTL/trait association can involve the evaluation of one marker at a time, two marker loci simultaneously, or the consideration of all possible marker loci at once. Typical of a one-marker comparison strategy is the use of the one-way analysis of variance (F test) for the analysis of BC progeny, and marker genotype means comparisons (*t* test) for BC and F_2 populations (Soller et al., 1976; Stuber et al., 1992). This approach ignores the potential recombination between a marker and a QTL, and thus will lead to an underestimation of QTL effects if the marker and QTL are not coincident (Edwards et al., 1987). A single marker approach may also incorporate a trait-based analysis in which individuals in the tails of a population distribution are sampled for marker frequencies (Lebowitz et al., 1987). In this case, those markers lying between the tails of the distribution and differing in frequency are assumed to be associated with the QTL affecting the trait.

Approaches which examine two marker loci at once incorporate interval mapping strategies using maximum likelihood for the analysis of single QTLs flanked by a pair of marker loci (Lander and Botstein, 1989; Paterson et al., 1991). The interval approach was developed to take advantage of additional information provided by linkage maps having a relatively high degree of genome saturation (i.e., spacing of markers every 5–20 cM) such as tomato and maize (Paterson et al., 1991; Doebly and Stec, 1991). The interval approach allows for the estimation of putative QTL effects at any location within a marker interval based on the means and variances observed in the marker classes and the recombination frequency between the markers bracketing a particular interval (Lander and Botstein, 1989). This approach is partially limited by its inability to test unlinked markers, and to accurately locate QTLs beyond the terminal markers of a given linkage group.

The consideration of all possible marker loci at once during QTL analysis is complex and involves the regression of trait expression on multiple marker locus values (Cowen, 1989; Rodolphe and Lefort, 1993; Stam, 1993). More recently, interval mapping and multiple regression have been integrated ("hybrid" approaches) to more accurately describe QTL/trait associations (Haley and Knott, 1992; Jansen, 1992, 1993; Jansen and Stam, 1994; Knapp, 1991; Knapp et al., 1990; Martinez and Curnow, 1992; Moreno-Gonzales, 1992; Zeng, 1993, 1994). Regardless of the mapping approach used, the success of MAS depends on the ability to detect QTLs and the consistency of QTLs over environments and generations (Dijkhuizen, 1994; Lande and Thompson, 1990; Shoemaker et al., 1994).

APPLICATION OF MARKERS

Genetic markers have been used effectively in genetic diversity analysis and germplasm organization [e.g., *Arachis* (Lanham et al., 1992), *Brassica* (dos Santos et al., 1994; Thormann et al., 1994), *Vaccinium* (Novy et al., 1994)]; in genetic similarity estimation as predictors of hybrid performance (Bernardo, 1994; Melchinger et al., 1990; Smith et al., 1990); in genetic map construction for the localization of loci conditioning simply inherited traits [e.g., *Pto* locus for resistance to *Pseudomonas syringae* pathovar tomato (*Pst*) (Carland and Staskawicz, 1993); *H1* gene for resistance to *Globodera rostochiensis* (Woll.) Behrens in potato (Gebhardt et al., 1993); *er-1* for resistance to powdery mildew in peas (*Pisum sativum* L.) (Timmerman et al., 1994); downy mildew resistance in lettuce (*Latuca sativa* L.) (Paran and Michelmore, 1993); photoperiod-sensitivity gene in rice (Mackill et al., 1993)]; and QTL analysis (Edwards et al., 1987; Table 3). Marker systems also provide the potential for map-based cloning of specific genes (Tanksley et al., 1995). Although theoretical appraisals of MAS have shown that it could be useful in plant improvement, the application of MAS has not been rigorously evaluated in many crop species (Lande and Thompson, 1990). To date, no cultivar developed through MAS has been publicly released.

Theoretical considerations and computer simulation

Theoretical investigations that probe the potential of MAS are of academic and practical importance. Although there are three general kinds of selection (stabilizing, directional, and disruptive) that could be used by plant breeders, directional selection is preferred because selected phenotypes are distinct from the initial population for economically important attributes. Truncation selection is the simplest type of directional selection. During truncation selection, a phenotypic value is identified as the lower selection limit (truncation point) and individuals are recovered whose phenotypic values are equal to or beyond this value. A prediction equation for response to truncation selection can be defined in terms of response to selection (R; difference in mean phenotype between the progeny generation and the previous generation), heritability (h^2), and the selection differential (S; difference in mean phenotype between the selected parents and the initial population mean) as: $R = h^2S$. Thus, realized heritability can be estimated in the first generation of purely phenotypic selection as: $h^2 = R/S$. Selection intensity (i) or selection differential is often expressed as units of standard deviations (s) in phenotypic value such that $i = S/s$.

Computer-based simulation can allow for tentative interpretation of relatively complex genetic comparisons that have not been previously possible (Edwards and Page, 1994; Lande and Thompson, 1990). Using simple relationships (e.g., $R = h^2S$) and theoretical assumptions of variance components in an initial population, Lande and Thompson (1990) proposed a computer-based simulation model for MAS to estimate genetic effects and gain from index selection. The model provides theoretical estimates for response from truncation selection for QTLs in an F_2 population using a 100-marker loci. The model derives selection indices that maximize the rate of improvement in quantitative characters under various methods of MAS. The model takes into account epistasis by combining multiplicative (multivariant) and classical additive approximations of gene action. Selection is based on an index that incorporates phenotypic and molecular information. The model uses the linkage disequilibrium between molecular marker loci and quantitative trait loci (QTLs) in populations created by a cross between two inbred lines.

Various strategies for plant improvement were tested by Lande and Thompson (1990) using computer simulations to characterize MAS and to provide expectations for phenotypic selection. Potential increases in breeding efficiency through MAS and the population size needed to attain such increases depends on the genetic parameters (i.e., heritability, the proportion of the additive genetic variance explained by the marker loci) and the selection method used. Gain from selection (ΔG) of quantitative traits based on estimated additive effects could be greater for MAS than for phenotypic selection. The relative worth of

Table 3. Estimated number of quantitative trait loci (QTLs) affecting the expression of traits in several crops.

Crop	Population	Trait	QTL (no.)	Range (%) of explained phenotypic σ^2[z]	LOD (range)	Total phenotypic σ^2 explained (%)	Reference
Common bean (*Phaseolus vulgaris* L.)	F_2:F_3	Nodule number	4	1.5–2.8	11.0–17.0	50.0	Nodari et al., 1993b
		Resistance to common blight	4	2.1–6.0	3.5–9.2	75.0	---
Corn (*Zea mays* L.)	F_3BC_1	Grain yield	8	5.6–14.4	6.33–10.86	---	Stuber et al., 1992
	F_3BC_2	Grain yield	6	6.2–18.0	3.16–9.73	---	---
	F_2	European corn borer resistance	7	2.3–9.1	3.4–15.7	38.0	Schon et al., 1993
		Plant height	3	5.7–12.9	10.3–34.1	63.0	---
	F_2:F_3[y]	Ear height	5	6.3–27.8	2.1–5.3	61.2	Veldboom et al., 1994
		Plant height	5	6.4–39.5	2.2–8.2	67.1	---
		GDD to anthesis[x]	6	2.1–5.9	6.2–33.6	62.9	---
		GDD to silk delay	2	2.8–3.2	15.8–17.5	30.9	---
		GDD to silk emergence	5	2.4–11.3	7.8–53.1	80.9	---
	F_2[w]	Number of cupules in single rank	6	4.1–24.6	2.6–11.0	---	Doebley and Stec, 1991
		Tendency of ear to shatter	6	4.3–41.7	2.4–18.6	---	---
		Hardness of outer glume	2	17.5–62.4	7.6–40.6	---	---
		Average length of internodes[v]	5	4.7–45.3	3.0–11.7	---	---
		Number of branches[u]	4	4.3–24.3	2.8–7.6	---	---
		Percent cupules lacking spikelet[t]	5	8.0–25.1	2.9–9.9	---	---
		Number of ears on lateral branch	7	6.3–24.5	2.8–12.9	---	---
		Percent male spikelets[t]	5	5.0–22.5	3.0–15.9	---	---
		Number of rows of cupules	6	5.0–36.0	2.8–15.9	---	---
Cowpea (*Vigna unguiculata* L.)	F_2	Seed weight	4	---	32.0–37	53.0	Fatokun et al., 1992
Potato (*Solanum tuberosum* L.)	BC	Type A trichome browning reaction	2	20.2–52.0	6.5–22.1	63.4	Bonierbale et al., 1994
		Type A trichome density	1	32.0	11.1	32.0	---
		Type A trichome polyphenol oxidase conc.	2	13.2–23.1	2.9–6.3	27.0	---
		Type B trichome sucrose ester levels	5	6.1–49.4	2.0–19.2	67.6	---
		Type B trichome density	2	8.6–35.4	2.9–14.2	38.1	---
	F_1	Late blight	19	---	---	---	Leonards-Schippers et al., 1994
	F_1	Tuber shape	1[s]	---	---	75.0[r]	van Eck et al., 1994
	F_1	Chip color	6	---	5.1–14.9	51.0	Douches and Freyre, 1994
Tomato (*Lycopersicon* spp.)	BC	Fruit mass	6	---	---	58.0	Paterson et al., 1988
		Soluble solids	4	---	---	44.0	---
		Fruit pH	5	---	---	48.0	---
	F_2	Insect resistance	3	---	---	---	Nienhuis et al., 1987
	F_2	Days to first true leaf	3	2.6–13.1	2.4–11.9	18.0	deVicente and Tanksley, 1993
		Days to first flower	7	3.5–10.2	2.2–8.0	43.0	---
		Plant height	9	3.1–8.4	2.2–8.2	42.0	---
		Total number of flower buds	10	2.8–34.0	2.2–34.1	61.0	---
		Number of internodes on primary stem	5	4.8–7.4	3.2–5.71	23.0	---
		Total number of internodes	9	3.7–12.8	2.8–10.8	52.0	---
		Number of well developed branches	8	3.1–9.5	2.7–7.5	53.0	---
		Total plant fresh weight	2	3.0–4.4	2.2–2.74	7.0	---
		Total plant dry weight	5	3.2–7.0	2.3–3.5	21.0	---
	F_2:F_3	Soluble solids	7	3.0–12.0	6.0–28.0	44.0	Paterson et al., 1991
		Fruit mass	11	2.3–21.5	4.0–42.0	72.0	---
		Fruit pH	9	2.4–6.1	4.2–28.0	34.0	---

[z]Given on a per locus basis.
[y]F_2 classified by F_3 families.
[x]Growing degree days.
[w]*Zea mays*; corn x teosinte.
[v]In primary lateral branches.
[u]In primary lateral inflorescence.
[t]Pedicellate spikelet.
[s]Multiple alleles were detected.
[r]Total genetic variance.

MAS is greatest for characters with low heritability when additive genetic variance is associated with the marker. More recently, Gimelfarb and Lande (1994b) have demonstrated that this same logic could be applied to nonadditive characters.

The simplifying assumptions of computer-based models, however, can lead to over- or under-estimations of R. For instance, fitness plays an integral part in response of individuals to selection. Heterotic advantage can be defined in terms of fitness. In many species, heterosis is pivotal to the expression of average fitness in a population. Estimates of R are only valid if an individual's "fitness" is interpreted as the probability that an individual is included among the group selected as parents in the next generation (Gimelfarb, 1989).

The potential utility of MAS in practical plant breeding programs is limited by: 1) the number of molecular marker loci required to detect all significant linkage associations; 2) population sizes required to detect QTLs for traits with low heritability; 3) the sampling errors associated with the weighting of indices when combining molecular marker loci; and 4) phenotypic information and the cost per

unit information gained (Edwards and Page, 1994; Lande and Thompson, 1990). Marker effectiveness (i.e., selection response) increases as the number of QTLs affecting a trait decreases (Edwards and Page, 1994). The effectiveness of MAS decreases as the linkage distance between markers and QTLs increases. Moreover, greater genetic gain can be made when flanking QTLs between two marker loci are used as compared to single markers, if single markers are loosely linked to a QTL. However, the use of flanking markers requires the characterization of twice as many markers as compared to selection using single markers. Thus, where dense maps are available, the value (cost/unit information) of flanking markers decreases as marker QTL associations increase.

The effectiveness of MAS is also determined by the relative linkage disequilibria between the marker loci and QTLs that condition trait expression (Lande and Thompson, 1990). Linkage disequilibria (between genetic markers and QTLs) is maximized by the mating germplasm of divergent origin. Such matings occur regularly during plant improvement (e.g., crossing between elite inbreds to begin genetic recombination for line development). Fixation of desirable trait loci in an elite background is the goal of breeding programs. Greater genetic gain is likely when fewer genes are involved in trait behavior because less recombination occurs (Edwards and Page, 1994). Having many QTLs exacerbates the problem of marker-QTL recombination, and thus the time required for fixation increases as the number of QTLs associated with a particular trait increases.

The multiple regression of phenotype on genetic markers can be used during MAS to provide a tactical assessment of gain from selection. Analysis of such relationships during MAS capitalizes on the linkage disequilibrium generated by the original mating of two inbred lines (Gimelfarb and Lande, 1994a). MAS can, therefore, be very effective during the early generations of population improvement where important linkages have not been eroded by recombination (Edwards and Page 1994; Lande and Thompson, 1990). Markers that contribute significantly to selection in initial generations should be re-evaluated each generation to determine their continued effectiveness (Gimelfarb and Lande, 1994a). This need can be costly and undermine the potential usefulness of MAS.

If a significant amount of the additive variance associated with a QTL can be accounted for by molecular markers, then MAS can increase breeding efficiency (Edwards and Page, 1994; Gimelfarb and Lande, 1994b). Likewise, the effectiveness of marker loci will be increased as the number of individuals in a population is increased, since a greater proportion of the additive genetic variance can potentially be explained. When trait heritability is low, population size must be relatively large (100–1000 individuals) to include unrelated individuals that detect additive variance associated with marker loci. Moreover, full- or half-sib populations are likely to require larger samples, depending on the degree of dominance associated with the trait.

Sample size is important when considering the potential loss of efficiency in MAS during protracted index selection (Gimelfarb and Lande, 1994a). Sampling errors can occur during model building as the relative weights of molecular and phenotypic information are estimated. MAS may only be cost efficient if phenotypic selection is made difficult due to large environmental effects and/or the number of loci affecting such traits is large (Lande and Thompson, 1990). However, increasing the number of markers that contribute to the selection index does not necessarily increase the effectiveness of MAS. The use of many markers may in fact result in a weaker response to selection.

Application

Simmonds (1979) has stated that "...plant breeding often not only generates benefits but is also attractive in having relatively low implementation costs..." This statement may no longer apply if MAS is rigorously applied to crop improvement. The costs of MAS can be high when compared to classical phenotypic selection, and the cost : benefit ratio may not be high enough to warrant use of MAS. The cost-benefit relationship can now be more critically evaluated in differing marker systems (Table 1). For instance, Lande (1992) indicated that the cost of scoring RFLPs is on the order of 100 to 1000 times as expensive as measuring standard phenotypes in most crops. Simulation experiments by Ragot and Hoisington (1993) indicated that the costs for employing RAPDs may be higher than for RFLPs as the number of individuals and markers in an experiment are increased. While RAPDs were found to be most cost/time efficient when sample sizes were small, RFLPs were more advantageous as sample sizes were increased. Darvasi and Soller (1994) considered experimental costs (i.e., number of individuals and marker spacing) that might be incurred during QTL analysis of a plant species possessing a genome size of 1000 cM. They concluded that the costs of MAS would be prohibitive since hundreds of individuals would be needed in a typical marker-QTL experiment, even if all possible markers could be used and the power of QTL detection (i.e., LOD) was high.

Although the costs associated with MAS are currently high, it has been shown to have potential utility for managing complex traits (Table 3). For instance, Stuber and Edwards (1986) recorded similar genetic gains when using either phenotypic selection or MAS (isozymes) for quantitative traits in maize. Similar results were found by Stromberg (as cited by Dudley, 1993). More recently, Stuber (1995) observed that significant genetic gain could be obtained during marker-assisted backcrossing in maize aimed at transferring targeted QTLs from T×303 into B73 and from Oh43 into Mo17. Although the yield of hybrids between the "enhanced" B73 and Mo17 exceeded that of control hybrids by more than 15%, no parallel assessment of the relative efficiency of classical breeding was included. Edwards and Johnson (1994) used MAS in two sweet corn populations (A and B) and found a positive response from selection (two cycles per year) for several traits. Elite lines were crossed to produce F_1 lines that were then subjected to either MAS (RFLPs) or phenotypic recurrent selection (PRS) for yield and quality traits. These original parents and the resultant populations (MAS and PSR) were crossed to two testers for replicated comparison at one location in one year. Although positive response to selection was observed for six of 11 traits in one population (A), differences between the average performances of the hybrids developed from MAS lines and parental hybrids were not significant. Likewise, the overall response to phenotypic or MAS selection was similar in a second population (B). The authors hypothesized that this lack of response to MAS was due to antagonistic effects of genome regions responsible for the yield and quality traits observed.

SUMMARY

Molecular markers and associated technologies can assist in map construction and the analysis of the molecular and genetic basis of quantitative and qualitative traits. Molecular markers that are tightly linked to economically important traits that are under the control of single genes have potential for immediate utility in plant improvement programs. Perhaps the most optimistic prospects for MAS is in disease resistance breeding, especially where several genes control resistance with complex interactions and where pyramiding of genes is desirable (Kelly, 1995; Schafer and Roelfs, 1985; Stavely et al., 1989).

Optimal use of QTL regions will, however, require a knowledge of their often complex epistatic interactions. Moreover, although MAS may have potential in population and inbred line development, it will likely have little or no effect in reducing the need for replicated field trials and testing. Optimal methods for mapping QTLs are still being debated and more sophisticated computer-aided analysis procedures are being developed. When QTLs and single genes are adequately mapped, they can be isolated biochemically (Tanskley et al., 1995; Young, 1990). Methods for their isolation (i.e., cloning) and characterization are also points of considerable discussion.

The effectiveness of any MAS procedure will depend on the accuracy of the phenotypic classification of trait expression and the degree of linkage between a marker(s) and traits of interest. Although MAS may increase gain from selection when compared to phenotypic selection, marker utility in plant improvement programs will ultimately be determined by cost/unit information (Edwards and Page, 1994). Clearly, laboratory costs associated with MAS applications are decreasing, and more effective and efficient molecular markers are being developed (Gu et al., 1995). This progress will make MAS more attractive and

will foster its prudent implementation as a tool for plant breeding. As a result of such changes, MAS might have potential for selection of characters such as yield components in agronomic and horticultural crops. Nevertheless, in horticultural crops, where many complex and highly integrated aesthetic, culinary, and organoleptic attributes are considered necessary for market acceptance, plant breeding expertise and decision making ability will clearly remain pivotal for genetic improvement.

Literature Cited

Akkaya, M.S., A.A. Bhagwat, and P.B. Cregan. 1992. Length polymorphisms of simple sequence repeat DNA in soybean. Genetics 132:1131–1139.

Albini, S.M. and G.H. Jones. 1987. Synoptonemal complex spreading in *Allium cepa* and *A. fistulosum*. I. The initiation and sequence pairing. Chromosoma 95:324–338.

Allard, R.W. 1956. Formulas and tables to facilitate the calculation of recombination values in heredity. Hilgardia 24:235–278.

Beckmann, J.S. and M. Soller. 1983. Restriction fragment length polymorphisms in genetic improvement: Methodologies, mapping and costs. Theor. Appl. Genet. 67:35–43.

Bernardo, R. 1994. Prediction of maize single-cross performance using RFLPs and information from related hybrids. Crop Sci. 34:20–25.

Bonierbale, M.W., R.L. Plaisted, O. Pineda, and S.D Tanksley. 1994. QTL analysis of trichome-mediated insect resistance in potato. Theor. Appl. Genet. 87:973–987.

Burr, B., F.A. Burr, K.H. Thompson, M.C. Albertsen, and C.W. Stuber. 1988. Gene mapping with recombinant inbreds in maize. Genetics 118:519–526.

Caetano-Anolles, D., B.J. Bassam, and P.M. Gresshoff. 1991. DNA amplification fingerprinting using very short arbitrary oligonucleotide primers. Biotechnology 9:553–557.

Carland, F.M. and B.J. Staskawicz. 1993. Genetic characterization of the *Pto* locus of tomato: Semi-dominance and cosegregation of resistance to *Pseudomonas syringae* pathovar tomato and sensitivity to the insecticide Fenthion. Mol. Gen. Genet. 239:17–27.

Causse, M.A., T.M. Fulton, Y.G. Cho, S.N. Ahn, J. Chunwongse, K. Wu, J. Xiao, Z. Yu, P.C. Ronald, S.E. Harrington, G. Second, S.R. McCouch, and S. Tanksley. 1994. Saturated molecular map of the rice genome based on an interspecific backcross population. Genetics 138:1251–1274.

Chakravarti, A., L.K. Lasher, and J.E. Reffer. 1991. A maximum likelihood method for estimating genome length using genetic linkage data. Genetics 128:175–182.

Churchill, G.A. and R.W. Doerge. 1994. Empirical threshold values for quantitative trait mapping. Genetics 138:963–971.

Cowen, N.M. 1989. Multiple regression analysis of RFLP data sets used in mapping QTLs, p. 113–116. In: T. Heletjaris and B. Burr (eds.). Development and applications of molecular markers to problems in plant genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Crow, J.F. 1990. Mapping functions. Genetics 125:669–671.

Darvasi, A. and M. Soller. 1994. Optimum spacing of genetic markers for determining linkage between marker loci and quantitative trait loci. Theor. Appl. Genet. 89:351–357.

de Vincente, M.C. and S.D. Tanksley. 1993. QTL analysis of transgressive segregation in an interspecific tomato cross. Genetics 134:585–596.

Diers, B.W., P. Klein, W.R. Fehr, and R.C. Shoemaker. 1992. RFLP analysis of soybean seed protein and oil content. Theor. Appl. Genet. 83:608–612.

Dijkhuizen, A. 1994. Application of restriction fragment length polymorphism for the assessment of genetic variability and study of quantitatively inherited traits in cucumber (*Cucumis sativus* L.). PhD Diss., Univ. of Wisconsin, Madison.

Doebly, J. and A. Stec. 1991. Genetic analysis of morphological differences between maize and teosinte. Genetics 129:285–295.

dos Santos, J.B., J. Nienhuis, P. Skroch, J. Tivang, and M.K. Slocum. 1994. Comparison of RAPD and RFLP genetic markers in determining genetic similarity among *Brassica oleracea* L. genotypes. Theor. Appl. Genet. 87:909–915.

Douches, D.S. and R. Freyre. 1994. Identification of genetic factors influencing chip color in diploid potato (Solanum spp.) Amer. Potato J. 71:581–590.

Dudley, J.W. 1993. Molecular markers in plant improvement: Manipulation of genes affecting quantitative traits. Crop Sci. 33:660–668.

Echt, C., S. Knapp, and B.H. Liu. 1992. Genome mapping with non-inbred crosses using GMendel 2.0. Maize Genet. Coop. Nwsl. 66:27–29.

Edwards, M. and L. Johnson. 1994. RFLPs for rapid recurrent selection. Proc. Symp. Analysis of Molecular Marker Data, p. 33–40. Aug. 1994. Corvallis, Ore.

Edwards, M.D. and N.J. Page. 1994. Evaluation of marker-assisted selection through computer simulation. Theor. Appl. Genet. 88:376–382.

Edwards, M.D., C.W. Stuber, and J.F. Wendel. 1987. Molecular facilitated investigations of quantitative trait loci in maize. I. Number, genomic distribution and types of gene action. Genetics 116:113–125.

Fatokun, C.A., D.I. Menancio-Havtea, D. Danesh, and N.D. Young. 1992. Evidence for orthologus seed weight genes in cowpea and mung bean based on RFLP mapping. Genetics 132:841–846.

Faure, S., J.L. Noyer, J.P. Horry, F. Bakry, C. Lanaud, and D. Gonzales de Leon. 1993. A molecular marker-based linkage map of diploid banana. Theor. Appl. Genet. 87:517–526.

Ferreira, M.E., P.H. Williams, and T.C. Osborn. 1994. RFLP mapping of *Brassica napus* using double haploid lines. Theor. Appl. Genet. 89:615–621.

Fisher, R.A. 1946. A system of scoring linkage data with special reference to pied factors in mice. Amer. Naturalist 80:568–578.

Fregeau, C.J. and R.M. Fourney. 1993. DNA typing with fluorescently tagged short tandem repeats: A sensitive and accurate approach to human identification. BioTechniques 15:100–119.

Gebhardt, C., D. Mugnieri, E. Reitter, F. Salamini, and E. Bonnel. 1993. Identification of RFLP markers closely linked to the H1 gene to conferring resistance to *Globodera rostochiensis* in potato. Theor. Appl. Genet. 85:541–544.

Gimelfarb, A. 1989. Genotypic variation maintained under stabilizing selection without mutations: Epistasis. Genetics 123:217–227.

Gimelfarb, A. and R. Lande. 1994a. Simulation of marker assisted selection in hybrid populations. Genetical Res. 63:39–47.

Gimelfarb, A. and R. Lande. 1994b. Simulation of marker assisted selection for non-additive traits. Genetical Res. 64:127–136.

Gu, W.K., N.F. Weeden, J. Yu, and D.H. Wallace. 1995. Large-scale, cost-effective screening of PCR products in marker-assisted selection applications. Theor. Appl. Genet. 91:465–470.

Gupta, M., Y.S. Chyi, J. Romero-Severson, and J.L. Owen. 1994. Amplification of DNA markers from evolutionarily diverse genomes using single primers of simple-sequence repeats. Theor. Appl. Genet. 89:998–1006.

Haldane, J.B.S. 1919. The combination of linkage values, and the calculation of distance between the loci of linked factors. J. Genet. 8:299–309.

Haldane, J.B.S. and C.A.B. Smith. 1947. A new estimate of the linkage between the genes for colour-blindness and haemophilia in man. Ann. Eugenics 14:10–31.

Haley, C.S. and S.A. Knot. 1992. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. Heredity 69:315–324.

Haley, S.D., L.K. Afanador, and J.D. Kelly. 1994. Selection for monogenic resistance traits with coupling- and repulsion-phase RAPD markers. Crop Sci. 34:1061–1066.

Hauge, B.M., S.M. Hanley, S. Cartinhour, J.M. Cherry, H.M. Goodman, M. Koorneef, P. Stam, C. Chang, S. Kempin, L. Medrano, and E.M. Meyerowitz. 1993. An integrated genetic/RFLP map of the *Arabidopsis thaliana* genome. Plant J. 3:745–754.

Jansen, R.C. 1992. A general mixture model for mapping quantitative trait loci by using molecular markers. Theor. Appl. Genet. 85:252–260.

Jansen, R.C. 1993. Interval mapping of multiple quantitative trait loci. Genetics 135:205–211.

Jansen, R.C. and P. Stam. 1994. High resolution of quantitative traits into multiple loci via interval mapping. Genetics 136:1447–1455.

Jarrell, D.C., M.L. Roose, S.N. Traugh, and R.S. Kupper. 1992. A genetic map of citrus based on the segregation of isozymes and RFLPs in an intergeneric cross. Theor. Appl. Genet. 84:49–56.

Jarvis, P., C. Lister, V. Szabo, and C. Dean. 1994. Integration of CAPs markers into the RFLP map generated using recombinant inbred lines of *Arabidopsis thaliana*. Plant Mol. Biol. 24:685–687.

Jiang, C. and Z. Zeng. 1995. Multiple trait analysis of genetic mapping for quantitative trait loci. Genetics 140:1111–1127.

Kelly, J.D. 1995. Use of random amplified polymorphic DNA markers in breeding from major gene resistance to plant pathogens. HortScience 30:461–465.

Kennard, W.C., K. Poetter, A. Dijkhuizen, V. Meglic, J. Staub, and M. Havey. 1994. Linkages among RFLP, RAPD, isozyme, disease resistance, and morphological markers in narrow and wide crosses of cucumber. Theor. Appl. Genet. 89:42–48.

Kleinhofs, A., A. Kilian, M.A. Shagai Maroof, R.M. Biyashev, P. Hayes, F.Q. Chen, N. Lapitan, A. Fenwick, T.K. Blake, V. Kanazin, E. Ananiev, L. Dahleen, D. Kudrna, J. Bollinger, S.J. Knapp, B. Liu, M. Sorrels, M. Heun, J.D. Franckowiak, D. Hoffman, R. Skadsen, and B.J. Steffenson. 1993. A molecular, isozyme and morphological map of the barley (*Hordeum vulgare*) genome. Theor. Appl. Genet. 86:705–712.

Knapp, S.J. 1991. Using molecular markers to map multiple quantitative trait loci: Models for backcross, recombinant inbred, and doubled haploid progeny. Theor. Appl. Genet. 81:333–338.

Knapp, S.J., W.C. Bridges, and D. Birkes. 1990. Mapping quantitative trait loci using molecular marker linkage maps. Theor. Appl. Genet. 79:583–592.

Konieczyn, A. and F.M. Ausubel. 1993. A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR-based markers. Plant J. 4:403–410.

Kosambi, D.D. 1944. The estimation of map distance from recombination values. Ann. Eugenics 12:172–175.

Kramer, H.H. and C.R. Burnham. 1947. Methods of combining linkage intensity values from backcross, F_2 and F_3 genetic data. Genetics 32:379–390.

Lande, R. 1992. Marker-assisted selection in relation to traditional methods of plant breeding, p. 437–451. In: H.T. Stalker and J.P. Murphy (eds.). Plant breeding in the 1990s. CAB Intl., Wallingford, U.K.

Lande, R. and R. Thompson. 1990. Efficiency of marker-assisted selection in the improvement of quantitative traits. Genetics 124:743–756.

Lander, E.S. and D. Botstein. 1986. Mapping complex genetic traits in humans: New method using a complex RFLP linkage map. Cold Spring Harbor Symposia on Quantitative Biology. 51:49–62.

Lander, E.S. and D. Botstein. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199.

Lander, E.S., P. Green, J. Abrahamson, A. Barlow, M.J. Daly, S.E. Lincoln, and L. Newberg. 1987. MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181.

Landry, B.S., N. Hubert, T. Etoh, J.J. Harada, and S.E. Lincoln. 1991. A genetic map for *Brassica napus* based on restriction fragment length polymorphism detected with expressed DNA sequences. Genome 34:543–552.

Lanham, P.G., S. Fennell, J.P. Moss, and W. Powell. 1992. Detection of polymorphic loci in *Arachis* germplasm using random amplified polymorphic DNAs. Genome 35:885–889.

Lark, K.G., K. Chase, F. Adler, L.M. Mansur, and J.H. Orf. 1995. Interactions between quantitative trait loci in soybean in which trait variation at one locus is conditional upon a specific allele. Proc. Natl. Acad. Sci. USA 92:4656–4660.

Lebowitz, R.J., M. Soller, and J.S. Beckmann. 1987. Trait-based analysis for the detection of linkage between marker loci and quantitative trait loci in crosses between inbred lines. Theor. Appl. Genet. 73:556–562.

Leonards-Schippers, C., W. Gieffers, R. Schäfer-Pregl, E. Ritter, S.J. Knapp, F. Salamini, and C. Gebhardt. 1994. Quantitative resistance to *Phytophthora infestans* in potato: A case study for QTL mapping in an allogamous plant species. Genetics 137:67–77.

Mackill, D.J., M.A. Salam, Z.Y. Wang, and S.D. Tanksley. 1993. A major photoperiod-sensitivity gene tagged with RFLP and isozyme markers in rice. Theor. Appl. Genet. 85:536–540.

Manly, K.F. and R.W. Elliot. 1991. MapManager, a microcomputer program for analysis of data from recombinant inbred strains. Mammalian Genome 1:123–126.

Markert, C.L. and F. Moller. 1959. Multiple forms of enzymes: Tissues, ontogenetic, and species specific patterns. Proc. Natl. Acad. Sci. USA 45:753–763.

Martinez, O. and R.N. Curnow. 1992. Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. Theor. Appl. Genet. 85:480–488.

Mather, K. 1938. Measurement of linkage in heredity. Methuen and Co., London.

Melchinger, A.E., M. Lee, K.R. Lamkey, A.R. Hallauer, and W.L. Woodmann. 1990. Genetic diversity for restriction fragment length polymorphism and heterosis for two diallel sets of maize inbreds. Theor. Appl. Genet. 80:488–496.

Michelmore, R.W., I. Paran, and R.V. Kesseli. 1991. Identification of markers linked to disease resistance genes by bulked segregant analysis: A rapid method to detect markers in specific genomic regions using segregating populations. Proc. Natl. Acad. Sci. USA 88:9828–9832.

Moreno-Gonzales, J. 1992. Genetic models to estimate additive and non-additive effects of marker-associated QTL using multiple regression techniques. Theor. Appl. Genet. 85:435–444.

Morgante, M., A. Rafalski, P. Biddle, S. Tingey, and A.M. Olivieri. 1994. Genetic mapping and variability of seven soybean simple sequence repeat loci. Genome 37:763–769.

Nienhuis, J., T. Helentjaris, M. Slocum, B. Ruzzero, and M. Schaefer. 1987. Restriction fragment length polymporphism analysis of loci associated with insect resistance in tomato. Crop Sci. 27:797–803.

Nodari, R.O., S.M. Tscai, R.L. Gilberston, and P. Gepts. 1993a. Towards an integrated linkage map of common bean. 2. Development of an RFLP-based linkage map. Theor. Appl. Genet. 85:513–520.

Nodari, R.O., S.M. Tsai, P. Guzman, R.L. Gilbertson, and P. Gepts. 1993b. Toward an integrated linkage map of common bean. 3. Mapping genetic factors controlling host–bacteria interaction. Genetics 134:341–350.

Nordheim, E.V., D.M. O'Malley, and S.C. Chow. 1984. On the performance of a likelihood ratio test for genetic linkage. Biometrics 40:785–790.

Novy, R.G., C. Kobak, J. Goffreda, and N. Vorsa. 1994. RAPDs identify varietal misclassification and regional divergence in cranberry [*Vaccinium macrocarpon* (Ait.) Pursh]. Theor. Appl. Genet. 88:1004–1010.

Olson, M., L. Hood, C. Cantor, and D. Botstein. 1989. A common language for physical mapping of the human genome. Science 245:1434–1435.

Owen, J.L. and C.M. Uyeda. 1991. Single primer amplification of avian genomic DNA detects polymorphic loci. Ann. Biotechnol. 2:107–122.

Paran, I., R. Kesseli, and R. Michelmore. 1991. Identification of RFLP and RAPD markers linked to downy mildew resistance genes in lettuce using near-isogenic lines. Genome 34:1021–1027.

Paran, I. and R.W. Michelmore. 1993. Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. Theor. Appl. Genet. 85:985–993.

Paterson, A.H., S. Damon, J.D. Hewitt, D. Zamir, H.D. Rabinowitch, S.E. Lincoln, E.S. Lander, and S.D. Tanksley. 1991. Mendelian factors underlying quantitative traits in tomato: Comparison across species, generations, and environments. Genetics 127:187–197.

Paterson, A.H., E.S. Lander, J.D. Hewitt, S. Peterson, S.E. Lincoln, and S.D. Tanksley. 1988. Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. Nature 335:721–726.

Philipp, U., P. Wehling, and S.G. Wricke. 1994. A linkage map of rye. Theor. Appl. Genet. 88:243–248.

Pillen, K., G. Streinruchen, G. Wricke, R.G. Herrmann, and C. Jung. 1992. A linkage map of sugar beet (*Beta vulgaris* L.). Theor. Appl. Genet. 84:129–135.

Rafalski, J.A. and S.V. Tingey. 1993. Genetic diagnostics in plant breeding: RAPDs, microsatellites and machines. Trends Genet. 9:275–279.

Ragot, M. and D.A. Hoisington. 1993. Molecular markers for plant breeding: Comparisons of RFLP and RAPD genotyping costs. Theor. Appl. Genet. 86:975–984.

Reinisch, A.J., J. Dong, C.L. Brubaker, D.M. Stelly, J.F. Wendel, and A.H. Paterson. 1994. A detailed RFLP map of cotton, *Gossypium hirsutum* x *Gossypium barbadense*: Chromosome organization and evolution in a disomic polyploid genome. Genetics 138:829–847.

Reiter, R.S., J.G.K. Williams, K.A. Feldman, A. Rafalski, S.V. Tingey, and P.A. Scolnik. 1992. Global and local genome mapping in *Arabidopsis thaliana* by using recombinant inbred lines and random amplied polymorphic DNAs. Proc. Natl. Acad. Sci. USA 89:1477–1481.

Risch, N. 1992. Genetic linkage: Interpreting LOD scores. Science 255:804.

Rodolphe, F. and M. Lefort. 1993. A multi-marker model for detecting chromosomal segments displaying QTL activity. Genetics 134:1277–1288.

Schafer, F.R. and A.P. Roelfs. 1985. Estimated relation between numbers of urediniospores of *Puccina graminis* f. sp. *tritici* and rates of occurrence of virulence. Phytopathology 75:749–750.

Schon, C.C., M. Lee, A.E. Melchinger, W.D. Guthrie, and W.L. Woodman. 1993. Mapping and characterization of quantitative trait loci affecting resistance against second generation European corn borer in maize with the aid of RFLPs. Heredity 70:648–659.

Shoemaker, R.C., L.L. Lorenzen, B.W. Diers, and T.C. Olson. 1994. Genome mapping and agriculture, p. 1–10. In: P.M. Gresshoff (ed.). Plant genome analysis. CRC Press, Boca Raton, Fla.

Shute, N.C.E. 1988. Power properties of the likelihood ratio test in an analysis of genetic diseases. Biometrics 44:951–958.

Simmonds, N.W. 1979. Principles of crop improvement. Longman Group, London.

Smith, O.S., J.S.C. Smith, S.L. Bowen, R.A. Tenborg, and S.J. Wall. 1990. Similarities among a group of elite maize inbreds as measured by pedigree. F_1 grain yield, heterosis and RFLPs. Theor. Appl. Genet. 80:833–840.

Soller, M., T. Brody, and A. Genizi. 1976. On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. Theor. Appl. Genet. 47:35–39.

Song, K.M., J.Y. Suzuki, M.K. Slocum, P.H. Williams, and T.C. Osborn. 1991. A linkage map of *Brassica rapa* (syn. *compestris*) based on restriction fragment length polymporphism loci. Theor. Appl. Genet. 82:296–304.

Southern, E.M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98:503–517.

Stam, P. 1993. Construction of integrated genetic linkage maps by means of a new computer package: JoinMap. Plant J. 3:739–744.

Stansfield, W.D. 1969. Theory and problems of genetics. Schaum's outline series. McGraw-Hill, New York.

Staub, J.E and L. Crubaugh. 1995. Selection for multiple lateral determinate cucumber genotypes. Cucurbit Gen. Coop. Rpt. 18:5–6.

Staub, J.E., L.J. Kuhns, B. May, and P. Grun. 1982. Stability of potato tuber isozymes under different storage regimes. J. Amer. Soc. Hort. Sci. 107:405–408.

Stavely, J.R., J.R. Steadman, and R.J. McMillian, Jr. 1989. New pathogenic variabiliy in *Uromyces appendiculatus* in North America. Plant Dis. 73:428–432.

Stuber, C. 1995. Yield improvement success using marker-facilitated QTL transfer between maize lines, p. 13. In: Plant Genome III. Abstr. Intl. Conf. Status of Plant Genome Res. 15–19 Jan. 1995, San Diego.

Stuber, C. and M.D. Edwards. 1986. Genotypic selection for improvement of quantitative traits in corn using molecular marker loci. Rpt. Ann. Corn Sorghum Res. Conf. Washington, D.C. 41:70–83.

Stuber, C.W., S.E. Lincoln, D.W. Wolf, T. Helentjaris, and E. Lander. 1992. Identification of genetic factors contributing to heterosis in a hybrid from two elite maize inbred lines using molecular markers. Genetics 132:823–839.

Suiter, K.A., J.F. Wendel, and J.S. Case. 1983. Linkage-1: A pascal computer program for the detection and analysis of genetic linkage. J. Hered. 74:203–204.

Swanson, C.P., T. Merz, and W.J. Young. 1990. Cytogentics: The chromosome in division, inheritance and evolution. Prentice Hall, New Delhi, India.

Tanksley, S.D., M.W. Ganal, and G.B. Martin. 1995. Chromosome landing: A paradigm for map-based gene cloning in plants with large genomes. Trends Genet. 11:63–68.

Tanksley, S.D., M.W. Ganal, J.P. Prince, M.C. de Vicente, M.W. Bonierbale, P. Broun, T.M. Fulton, J.J. Giovannoni, S. Grandillo, G.B. Martin, R. Messeguer, J.C. Miller, L. Miller, A.H. Paterson, O. Pineda, M.S. Roder, R.A. Wing, W. Wu, and N.D. Young. 1992. High density molecular maps of the tomato and potato genomes. Genetics 132:1141–1160.

Tanksley, S.D., J. Miller, A. Paterson, and R. Bernatsky. 1988. Molecular mapping of plant chromosomes, p. 157–173. In: J.P. Gustafson and R. Appels (eds.). Chromosome structure and function: Impact of new concepts. Plenum Press, New York.

Teutonico, R.A. and T.C. Osborn. 1994. Mapping of RFLP and qualitative trait loci in *Brassica rapa* and comparison to the linkage map of *B. napus*, *B. oleracea* and *Arabidopsis thaliana*. Theor. Appl. Genet. 89:885–894.

Thormann, C.E., M.E. Ferreira, L.E. Camargo, J.G. Tivang, and T.C. Osborn. 1994. Comparison of RFLP and RAPD markers to estimating genetic relationships within and among cruciferous species. Theor. Appl. Genet. 88:973–980.

Timmerman, G.M., T.J. Frew, N.F. Weeden, A.L. Miller, and D.S. Goulden. 1994. Linkage analysis of *er-1*, a recessive *Pisum sativum* gene for resistance to powdery mildew fungus (*Erysiphe pisi* D.C.). Theor. Appl. Genet. 88:1050–1055.

Vallejos, C.E., N.S. Sakiyama, and C.D. Chase. 1992. Molecular marker-based linkage map in *Phaseolus vulgaris* L. Genetics 131:733–740.

van Eck, H.J., J.M.E. Jacobs, P. Stam, J. Ton, W.J. Stickema, and E. Jacobsen. 1994. Multiple alleles for tuber shape in diploid potato detected by qualitative and quantitative genetic analysis using RFLPs. Genetics 137:303–309.

Veldboom, L.R., M. Lee, and W.L. Woodman. 1994. Molecular marker facilitated studies in an elite maize population: 1. Linkage analysis and determination of QTLs for morphological traits. Theor. Appl. Genet. 88:7–16.

Webb, D.M., S.J. Knapp, and L.A. Tagliani. 1992. Restriction fragment length polymorphism and allozyme linkage map of *Cuphea lanceolata*. Theor. Appl. Genet. 83:528–532.

Weeden, N.F., M. Timmerman, M. Hermmat, B.E. Kneen, and M.A. Lodhi. 1992. Inheritance and reliability of RAPD markers. In: J. Nienhuis (ed.). Proc. Symp. Applications of RAPD Technology to Plant Breeding, p. 12–17. Nov. 1992, Minneapolis, Minn.

Welsh, J. and M. McClelland. 1990. Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res. 18:7213–7218.

Whitkus, R., J. Doebley, and M. Lee. 1992. Comparative genome mapping of sorghum and maize. Genetics 132:1119–1130.

Williams, J.G.K., A.R. Kubelik, K.J. Livak, J.A. Rafalski, and S.V. Tingey. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res. 18:6531–6535.

Xu, G.W., C.W. Magill, K.F. Schertz, and G.E. Hart. 1994. A RFLP linkage map of *sorghum bicolor* (L.) Moendr. Theor. Appl. Genet. 89:139–145.

Young, N. and S.D. Tanksley. 1989. RFLP analysis of the size of chromosomal segments retained around the *Tm-2* locus of tomato during backcross breeding. Theor. Appl. Genet. 77:353–359.

Young, N.D. 1990. Potential applications of map-based cloning to plant pathology. Physiol. Mol. Plant Pathol. 37:81–94.

Zabeau, M. and P. Vos. 1993. Selective restriction fragment amplification: A general method for DNA fingerprints. European Patent Application. Publ. 0534858A1.

Zamir, D. and Y. Tadmor. 1986. Unequal segregation of nuclear genes in plants. Bot. Gaz. 147:355–358.

Zeng, Z.B. 1993. Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. Proc. Natl. Acad. Sci. USA 90:10972–10976.

Zeng, Z.B. 1994. Precision mapping of quantitative trait loci. Genetics 136:1457–1468.

Zietkiewicz, E., A. Rafalski, and D. Labuda. 1994. Genome fingerprinting by simple sequence repeat (SSR)–anchored polymerase chain-reaction amplification. Genomics 20:176–183.

Chapter 9 in: **Which DNA Marker for Which Purpose?** Final Compendium of the Research Project **Development, optimisation and validation of molecular tools for assessment of biodiversity in forest trees** in the European Union DGXII Biotechnology FW IV Research Programme **Molecular Tools for Biodiversity**. Gillet, E.M. (ed.). 1999. URL http://webdoc.sub.gwdg.de/ebook/y/1999/whichmarker/index.htm

Chloroplast microsatellites for analysis of the geographic distribution of diversity in conifer species

M. Anzidei[1], A. Madaghiele[1], C. Sperisen[2], B. Ziegenhagen[3], G.G. Vendramin[1]*

[1] Istituto Miglioramento Genetico Piante Forestali, Consiglio Nazionale delle Ricerche, Via Atto Vannucci 13, 50134 Firenze, Italy

[2] Eidgenössische Forschungsanstalt für Wald, Schnee und Landschaft, Zürcherstrasse 111, 8903 Birmensdorf, Switzerland

[3] Institut für Forstgenetik und Forstpflanzenzüchtung, Bunderforschungsanstalt für Forst- und Holzwirtschaft, Sieker Landstrasse 2, 22927 Grosshansdorf, Germany

*Corresponding author: Email: vendramin@imgpf.fi.cnr.it

Introduction

In recent years great efforts have been put on the identification of highly informative molecular markers for screening diversity in forest tree species. The identification of efficient methods, in terms of costs and time, for accurate analyses of patterns of diversity in forest trees is extremely relevant, because these organisms typically show a high level of variability, so that sampling of a large number of populations and individuals is required for each study. Microsatellites are tandem repeats characterised by short motifs (1 to 6 bp), a low degree of repetition (5 to 100 repeat units) and a randomly dispersed distribution of about 10^4 to 10^5 microsatellite regions per genome (Tautz, 1993). They display a presumably selectively neutral behaviour, show a co-dominant inheritance that allows discrimination of homo- and heterozygotic states, and occur frequently and evenly distributed throughout the genome. Their high degree of length polymorphism, which is due to different numbers of repeats within the microsatellite regions, can be easily and reproducibly detected via the polymerase chain reaction (PCR). Their main applications are in genome mapping and in population analysis, but microsatellites are also useful for taxonomy, parentage analysis, identification of individuals in forensic studies, and human cancer diagnostics.

Microsatellites are not limited to the nuclear genome. They occur in chloroplasts and also in the mitochondrial genome, as found by Soranzo *et al.* (1999) as a repetition of G/C. This study reviews work that was done in the field of chloroplast microsatellites in conifer species.

Chloroplast microsatellites (cpSSR)

The availability of the entire chloroplast sequence of the Japanese pine species *Pinus thunbergii* (Wakasugi *et al.*, 1994) allowed the identification of cpSSR (chloroplast single sequence repeats). These microsatellites consist of repetitions of a single nucleotide (19 A/T and 1 G/C) (Powell *et al.*, 1995;

Vendramin *et al.*, 1996). Primers for the amplification of chloroplast microsatellites were also recently designed for angiosperms (Weising and Gardner, 1999).

Considering that the chloroplast genome does not recombine due to its paternal inheritance in conifers (*e.g.* for cp-microsatellites: Cato and Richardson, 1996 ; Vendramin and Ziegenhagen, 1997; Sperisen *et al.*, 1998) and maternal inheritance in angiosperms (*e.g.* for PCR/RFLP polymorphism in oaks: Dumolin *et al.*, 1995), cpSSR variants accumulate in a uniparental chloroplast lineage and can thus provide information about the history of populations. Microsatellite variants are supposed to be generated in a stepwise manner by addition or deletion of single repeat units. Under such a stepwise mutation model (Valdes *et al.*, 1993), microsatellite variants with small repeat length differences are more closely related than alleles with larger length differences, and consequently it can be considered that the process of mutation has a "memory" (Jarne and Lagoda, 1996). Computer simulations have produced linear relationships between genetic distances based on the size differences of the SSR alleles and the time of divergence (Slatkin, 1995; Goldstein *et al.*, 1995).

Methods for the detection of chloroplast microsatellite polymorphism

Standard methods were optimised for the characterisation and the screening of chloroplast microsatellites in conifers. The main steps of the procedures for the amplification and detection of length polymorphism can be summarised as follows (for details see Vendramin *et al.*, 1996).

PCR conditions

PCR amplifications were carried out using a Perkin Elmer model 9600 thermal cycler in a total volume of 25 µl containing 0.2 mM of each dNTPs, 2.5 mM of $MgCl_2$, 0.2 µM of each primer, 10x reaction buffer (Pharmacia), 25 ng of template DNA and 1 unit of Pharmacia Taq polymerase, with the following profile: 5 min denaturation at 95°C, 5 min at 80°C enzyme addition, followed by 25 cycles of 1 min denaturation at 94°C, 1 min annealing at 55°C and 1 min extension at 72°C, with a final extension step of 72°C for 8 min. Amplification reactions were automatically prepared using a robotic workstation Biomek 2000 (Beckman Instruments). One of the two PCR primers in each reaction was 5' fluorescine-labelled.

Sizing and sequencing of amplification products

Several pairs of primers for the amplification of chloroplast microsatellites were designed in order to obtain fragments having different size ranges, thus allowing multiplexing by size range. Two or three microsatellite-producing fragments in different size ranges were simultaneously loaded together with internal molecular weight standards on a 6%, 20 cm long denaturating 7M urea, 0.6x Tris-Borate-EDTA polyacrilamide gel (Pharmacia) and run on an ALF automatic sequencer (Pharmacia) at 35 Watt constant power for approximately 80 minutes. The same gel was loaded twice. External molecular weight standards as well as internal standards (50, 100, 150 and 200 bp) were used in conjunction with the Fragment Manager version 1.2 conversion software (Pharmacia) to size the amplified fragments. An example of sizing of chloroplast microsatellites using the automatic ALF Pharmacia sequencer is reported in Figure 1.

Selected polymorphic amplification products were sequenced in order to confirm the presence of the microsatellite regions in the amplified fragments as well as to verify that length variation was due to a different number of repeats within the microsatellite regions and not to mutations in the flanking regions. For this purpose amplified fragments were cloned into a PCR II plasmid vector (*Invitrogen*) and then sequenced using the ALF automatic sequencer (Pharmacia). The clones were sequenced from both

ends, using M-13 universal fluorescent-labelled primers and T7 DNA polymerase (Pharmacia). The sequences were run on 6% polyacrilamide, 7M urea, 0.6x Tris-Borate-EDTA gels at 1500V, 38mA and 34 Watt for 3 hours. Two clones from each cloning experiment were sequenced. An example of the sequencing of a chloroplast microsatellite region in *Abies alba* is reported in Figure 2.

Molecular organisation of chloroplast microsatellites in two conifer species

Abies alba Mill. (silver fir)

Sequencing was done for those amplification products revealing polymorphism between parents involved in controlled crosses and exhibiting sufficient polymorphism (Pt 30204 and Pt 71936, codes as in Vendramin *et al.*, 1996). Figure 2a gives the sequences for microsatellite locus Pt 30204 of three *A. alba* individuals (A,B,C) and one *A. pinsapo* individual. Figure 2b depicts the sequences for microsatellite locus Pt 71936 of *A. alba* individual C and the *A. pinsapo* individual. Moreover, in Figures 2a and b, alignment of sequences of the two microsatellite loci was done for the parental *Abies* individuals and *Pinus thunbergii.*

Figures 2a and b clearly reveal that the amplified loci contain simple sequence repeats and that their length polymorphism is due to a variable number of repeat units, thus confirming these loci to be chloroplast microsatellites. In the completely sequenced chloroplast genome of *Pinus thunbergii* (see above) the microsatellite locus Pt 30204 is characterised by a jointly occurring C and T mononucleotide stretch $(C)_{10}$ x $(T)_{12}$. In the investigated individuals of *Abies* this locus turns out to be composed of variable numbers of three mononucleotide repeats (Figure 2a). Microsatellite locus Pt 71936 contains a mononucleotide repeat characterised by a variable number of T for all three species *Abies alba, A. pinsapo* and *P. thunbergii* (Figure 2b).

For locus Pt 30204, alignment of sequences exhibits a striking heterogeneity of molecular organisation within the species *A. alba* (Figure 2a). Individual A is different from conspecific individuals B and C by insertion/deletions within the mononucleotide stretches, an addition of cytosine repeat units, and also by heterologous sequences of the interspersed non-coding sequences. By this organisation, the *A. alba* individual resembles more that of *P. thunbergii* (93% homology) than that of the conspecific individuals B and C (*e.g.* 79% homology to B), while B and C only differ by an insertion/deletion of 1 adenosine in the repetitive stretch. Also striking is the finding that for this microsatellite locus the *A. pinsapo* individual is more homologous to the *A. abies* individuals B and C than is A to either B or C.

The molecular organisation of SSR locus Pt 71936 seems to be more conserved throughout all sequences (Figure 2b). Size variations are more or less restricted to variable numbers of repeat units.

Sequence analysis of the microsatellite Pt30205 revealed that in many cases individuals sharing the same size are also identical in sequence. Nevertheless, individuals carrying the same size variants are not always characterised by the same microsatellite sequence, as was discovered by analysing individuals belonging to different populations as well as individuals of the same population. This indicates that size variation may underestimate the underlying sequence variation (Ziegenhagen *et al.*, in preparation).

Picea abies (L.) Karst. (Norway spruce)

Sequence analysis of amplified fragments revealed the occurrence of microsatellites and showed that size variation at the same microsatellite locus was due to variation in the copy number of SSRs (Sperisen *et al.*, 1998). The three microsatellites (Pt 26081, Pt 63718 and Pt 71936, codes as in

Vendramin *et al.*, 1996) consisted of A/T mononucleotide repeats. Alignment of the *Pinus thunbergii* and Norway spruce sequences showed a very high degree of homology. Size variation at chloroplast microsatellite loci Pt 71936 and Pt 63718 was restricted to differences in the copy number of SSRs, with the flanking sequences being identical. The sequences at microsatellites locus Pt 26081 showed variation in the copy number of SSRs and three insertions/deletions and a nucleotide substitution in the regions flanking the SSRs. Moreover, microsatellite locus Pt 71936 of Norway spruce revealed an organisation similar to that of the same microsatellite in silver fir. Both sequences contained a SSR and differed by one insertion/deletion and four nucleotide substitutions in the regions flanking the SSRs.

Inheritance of chloroplast microsatellites

Inheritance of chloroplast microsatellites was tested in Norway spruce (*Picea abies* K.) by Sperisen *et al.* (1998) and in silver fir (*Abies alba* Mill.) by Vendramin and Ziegenhagen (1997).

The mode of inheritance of the chloroplast microsatellites in Norway spruce was analysed in the progenies of seven controlled crosses including a reciprocal cross (Sperisen *et al.*, 1998). The progenies of all crosses exclusively showed the size variant found in the male parent, thus indicating that the three chloroplast microsatellites are paternally inherited (Table 1). The absence of the size variant of the female parent in the embryos appeared to exclude the occurrence of heteroplasmy as a result of maternal leakage.

SSR*	Cross	Female parent size variants (bp)	Male parent size variants (bp)	Size variants (bp) of progenies (N)
Pt 26081	1895 x 5444	116	115	115 (24)
	5443 x 39	115	116	116 (24)
	5460 x 2037	116	111	111 (24)
Pt 63718	1641 x 5468	99	102	102 (24)
	1895 x 5444	100	96	96 (24)
	5443 x 39	100	96	96 (24)
	reciprocal cross	96	100	100 (24)
	reciprocal cross	100	96	96 (24)
	5451 x 5444	100	96	96 (24)
Pt 71936	1895 x 5444	145	146	146 (24)

Table 1: Transmission of chloroplast microsatellites in Norway spruce as studied in intraspecific crosses (from Sperisen *et al.*, 1999). bp = base pairs; N = number of seeds analysed; * codes refer to Vendramin *et al.* (1996).

Using primer pairs derived from chloroplast simple sequence repeats of *Pinus thunbergii,* two polymorphic SSR loci were identified and sequence-characterised in the genus *Abies* (Vendramin and Ziegenhagen, 1997). PCR products exhibited considerable length variation among six different *Abies* species and within *A. alba*. A total of 75 F1 progeny of both an interspecific and an intraspecific reciprocal cross confirmed the two SSRs to be stably inherited and to follow predominantly paternal

inheritance (Table 2). When in addition to the embryo of each seed also the primary haploid endosperm (megagametophyte) was analysed, the size variant of the seed parent predominantly occurred, thus giving evidence for the elimination of the maternal plastid only in the egg cell or in the pro-embryos. In Table 3, results are given for PCR amplification of genomic DNA from nine *Abies* individuals from six different species. From the 11 tested primer pairs, two generated amplification products (Pt 30204 and Pt 71936) which showed considerable length variation among *A. alba* individuals but also among the six investigated *Abies* species (Table 3, shaded rows). Seven size variants of PCR products were observed when using primer pair Pt 30204 and eight size variants when using Pt 71936. One primer pair (Pt 15169) exhibited size variation only for the *A. cephalonica* individual. Six primer pairs produced amplification products, each of identical length for all individuals under study. One primer pair does very poor amplification, and the remaining pair failed to generate any amplification product at all.

Paternal inheritance of chloroplast microsatellites was also tested and confirmed in interspecific crosses between *Pinus halepensis* Mill. and *Pinus brutia* Ten. (Anzidei *et al.*, in preparation).

SSR	Maternal parent	Paternal parent	Seedling progeny	Seed progeny	
				Embryos	Megagametophytes
	Size variant	Size variant	Size variant (n of N)	Size variant (n of N)	Size variant (n of N)
Pt 71936 Pt 30204	***A. alba C*** **147** **136**	***A. pinsapo*** **156** **139**	156 (6 of 6) 139 (6 of 6)		
Pt 30204	***A. alba* A** **147**	***A. alba* B** **37**		137 (30 of 30)	47 (18 of 18) and 137 (2 of 18)
Pt 30204	***A. alba* B** **37**	***A. alba* A** **47**		147 (33 of 33) and 137 (1 of 33)	137 (23 of 23)

Table 2: Transmission of cp microsatellites (Pt 30204 and Pt 71936) in *Abies* as studied in an interspecific and in an intraspecific reciprocal cross (from Vendramin and Ziegenhagen, 1997). n = number of observed size variants, N = number of investigated individuals, embryos or megagametophytes.

Code*	**Primer sequences 5' - 3'** **sense** **antisense**	**Abies individuals No. 1 - 9** **size of PCR products [bp]**								
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Pt 9383	AGA ATA AAC TGA CGT AGA TGC CA AAT TTT CAA TTC CTT TCT TTC TCC	no amplification								
Pt 15169	CTT GGA TGG AAT AGC AGC C GGA AGG GCA TTA AGG TCA TTA	131	131	131	131	131	131	131	131	130
Pt 26081	CCC GTA TCC AGA TAT ACT TCC A TGG TTT GAT TCA TTC GTT CAT	112	112	112	112	112	112	112	112	112

Pt 30204	TCA TAG CGG AAG ATC CTC TTT CGG ATT GAT CCT AAC CAT ACC	147	137	136	149	140	149	139	149	148
Pt 36480	TTT TGG CTT ACA AAA TAA AAG AGG AAA TTC CTA AAG AAG GAA GAG CA	149	149	149	149	149	149	149	149	149
Pt 41093	TCC CGA AAA TAC TAA AAA AGC A CTC ATT GTT GAA CTC ATC GAG A	poor amplification								
Pt 48210	CGA GAT TGA TCC GAT ACC AG GAG AGA ACT CTC GAA TTT TTC G	89	89	89	89	89	89	89	89	89
Pt 71936	TTC ATT GGA AAT ACA CTA GCC C AAA ACC GTA CAT GAG ATT CCC	149	149	147	148	151	150	156	152	150
Pt 79951	CTT TTG TTT TTC AAC AAT TGC A ACA TCT ATC TCC CAT ATC GGC	138	138	138	138	138	138	138	138	138
Pt 87268	GCC AGG GAA AAT CGT AGG AGA CGA TTA GAC ATC CAA CCC	171	171	171	171	171	171	171	171	171
Pt 110048	TAA GGG GAC TAG AGC AGG CTA TTC GAT ATT GAA CCT TGG ACA	90	90	90	90	90	90	90	90	90

Table 3: Size of PCR products obtained amplifying genomic DNA of six *Abies* species with 11 primer pairs matching flanking regions of simple sequence repeats (SSRs) in cpDNA of *Pinus thunbergii* (Wakasugi *et al.* 1994) (from Vendramin and Ziegenhagen, 1997). * denotes the position of the 5' base of sense primer in the published *P. thunbergii* cpDNA sequence. Individuals No.1: *A. alba* A; 2: *A. alba* B; 3: *A. alba* C; 4: *A. nordmanniana*; 5: *A. cilica*; 6: *A. numidica*; 7: *A. pinsapo*; 8: *A. nordmanniana*; 9: *A. cephalonica*

Universality of chloroplast microsatellite markers

The identification of microsatellite regions is a very expensive and time-consuming process, which generally requires the construction and screening of a genomic library (efficient protocols for the enrichment in microsatellites are available, *e.g.* Edwards *et al.*, 1996). Therefore great efforts are necessary for sequencing and for the optimisation of the markers. Generally not more than 25% of the identified microsatellites are single locus Mendelian markers. A possible strategy to try to increase the efficiency of the identification of microsatellite regions, besides the construction of enriched libraries, may be represented by the possibility to transfer microsatellite markers developed in one species to others in order to reduce the costs of their development.

In the case of nuclear microsatellites, this strategy has been proved to be not very efficient. Echt *et al.* (1998), for example, using SSR primer pairs from *Pinus strobus* and *Pinus radiata*, found that while primers for monomorphic loci could amplify loci from a wide range of species, the primers for

informative dinucleotide repeat loci could only amplify loci from species within the same subgenus. Primers for the amplification of nuclear microsatellites in *Pinus halepensis* also work in the closely related species *Pinus brutia* (Keys *et al.*, 1999). Twenty-five primer pairs developed for *Pinus halepensis* were also tested in *Pinus pinaster*, but only one of them produced polymorphic amplification products that showed only a single band per haploid genome; the remaining 24 pairs produced either no amplification product, single monomorphic bands, or multiband patterns (Mariette *et al.*, 1999).

For the chloroplast genome, on the contrary, the high level of DNA sequence conservation, including the arrangements of genes and intergenic sequences in conifers as well as in angiosperms, confer to the cpSSR markers a very high degree of "universality". Thus, primers designed on the basis of the sequences of the chloroplast genome of *Pinus thunbergii* also worked in many other conifers (Powell *et al.*, 1995; Vendramin *et al.*, 1996) as well as in angiosperms (Cato and Richardson, 1996). The high degree of conservation of sequences in the chloroplast genome of conifers was also confirmed by studies performed by Vendramin *et al.* (1996) in *Pinus leucodermis*, Vendramin and Ziegenhagen (1997) in *Abies alba* and Sperisen *et al.* (1998) in *Picea abies*. These primers have been used with success in 110 different conifer species belonging to different taxonomic classifications, in particular to the *Pinaceae, Cupressaceae* and *Taxodiaceae* (Vendramin *et al.*, unpublished data), thus dramatically reducing the cost of development of these markers. Sequencing data in *Picea abies, Abies alba, Pinus halepensis, Pinus brutia, Pinus pinaster, Pinus pinea,* and *Pinus cembra* always confirmed the presence of the microsatellite region in the amplified fragments. Analyses are also in progress to verify the presence of cpSSRs in the amplified fragments of *Cupressus sempervirens* and *Taxus baccata.* The detection of a typical 1bp variation of the amplified fragments of different individuals of C*upressus sempervirens* seems to be evidence of the presence of cpSSR regions also in the *Cupressaceae.*

Equipment and cost

In principle, only basic molecular biology facilities are necessary for the analysis of chloroplast microsatellite polymorphism, such as a PCR thermal cycler and a system for vertical gel electrophoresis. However, considering that chloroplast microsatellites are repetitions of a single nucleotide and that therefore single base length variation must be detected, a system with high resolution (sequencing gel apparatus) should be available. High accuracy for sizing the amplified fragments can be obtained using internal and external standards of known molecular weight (see Figure 1).

The efficiency of the screening can, however, be increased considerably through the automation of the procedures: multiplex of PCR reaction and/or multiple loading of a single gel (at least twice) as well as the use of automated DNA sequencing apparatus (which does not require radioactive labelling) with appropriate software can speed up and automate the genotyping to a considerable extent. The use of a robotic workstation equipped with a multi-channel pipettor allows the preparation of 96 amplification reactions in about 20 minutes. The simultaneous use of a Beckman 2000 robotic workstation and of an ALF automatic sequencer with the software Fragments Manager version 1.2 allows the amplification and sizing of 300 to 500 samples with three chloroplast microsatellites per day per operator.

Case studies

Geographic distribution of the diversity

Picea abies

Three chloroplast microsatellites (cpSSRs), previously sequence-characterised and for which paternal inheritance was tested and confirmed (Sperisen *et al.*, 1998), were used to assess their usefulness as

informative markers for phylogeographic studies in Norway spruce (*Picea abies*) and to detect spatial genetic differentiation related to the possible recolonization processes in the post-glacial period (Vendramin *et al.*, 2000). Ninety-seven populations were included in the survey. Some 8, 7 and 6 different size variants for the three cpSSRs, respectively, were scored by analysing 1,105 individuals. The above 21 variants combined into 41 different haplotypes. The distribution of some haplotypes showed a clear geographic structure and seems to be related to the existence of different refugia during the last glacial period (Figure 3). The haplotype 03 (*116/96/144* as variant size at the three loci analysed, respectively) was present only in Scandinavia and northeastern and southeastern Europe, while the haplotype 04 (*116/100/143)* was detected only in the western part of the natural range (Central Europe and the Alps - Figure 3). The analysis of chloroplast SSR variation revealed the presence of two main gene pools (Sarmathic-Baltic and Alpine-Centre European) and a relatively low degree of differentiation (R_{st} (Slatkin, 1995) of about 10%), characteristic of tree species with large distribution and probably influenced by intensive human impact on this species. No evidence for the existence of additional gene pools (*e.g.* from Balkan and Carpathian glacial refugia) were obtained, though the existence of genetic discontinuity within the species' European range was observed (Figure 4).

Geostatistics was applied to the chloroplast simple sequence repeats (cpSSR) haplotype-frequency data from the 95 European Norway spruce populations (Bucci and Vendramin, 2000) to provide preliminary evidence about the following issues: 1) delineation of genetically homogeneous regions ('genetic zones'); 2) prediction of their haplotype frequencies and definition of related criteria to be applied for provenance identification and certification of seedlots; 3) construction of a continental-scale 'availability map' of the intraspecific biodiversity for Norway spruce. Direct evidence for large-scale geographic structure over the European natural range was obtained, detecting both geographic clines and stationary patterns. The increase of the mean genetic divergence by geographic distances (up to about 1,800 km) provided a strong hint that geographic distance is a major factor for differentiation in Norway spruce. Haplotype-frequency surfaces were obtained by applying ordinary kriging to sampling frequency data. Cluster analysis carried out on haplotype-frequency surfaces revealed a fair discrimination among 16 genetic zones (Figure 5). Dendrogram analysis carried out on the predicted mean haplotype frequency confirmed a fairly good separation of the detected genetic zones (Figure 6).

Application of geostatistical analysis to the large amount of genetic data is a promising tool for the analysis of complex geographic patterns aimed to reconcile appropriate conservation strategies and breeding exploitation of genetic resources.

Abies alba

Based on two polymorphic chloroplast microsatellites that had been previously identified and sequence-characterised in the genus *Abies (*Vendramin and Ziegenhagen, 1997*)*, genetic variation was studied in a total of 714 individuals from 17 European silver fir (*Abies alba* Mill.) populations distributed all over the natural range (Vendramin *et al.*, 1999). Eight and 18 different length variants at the two loci, which combined into 90 different haplotypes, were detected (Figure 7). Genetic distances between most populations as measured by d_0 (Gregorius 1984) were high and significant. There is also evidence for spatial organisation of the distribution of haplotypes, as shown by permutation tests, which demonstrate that genetic distances increase with spatial distances (Figure 8). Large heterogeneity in diversity across populations was observed, as measured by the number of haplotypes, by the unbiased index of diversity of Nei (1973) as well as by allelic richness. Furthermore, there is good congruence in the levels of allelic richness of the two loci across populations. The present organisation of levels of allelic richness across the range of the species is likely to have been shaped by the distribution of refugia during the last glaciation and the subsequent recolonization processes. Mainly, processes of genetic drift due to bottleneck and effects of isolation could be inferred from the obtained data.

Pinus halepensis (Aleppo pine) and *Pinus brutia* (Brutian pine)

Nine chloroplast, paternally inherited SSR markers were used to describe genetic variation of the two closely related species belonging to the *halepensis* complex (*P. halepensis* Mill. and *P. brutia* ten.). They reveal a large polymorphism both within and among populations (Bucci *et al.*, 1998). The high level of among-population genetic divergence (R_{st}, Slatkin, 1995) found for *P. halepensis* and the low within-population haplotypic diversity (except for Greek and Southern Italian stands) (S_w, Slatkin, 1995) are consistent with the hypothesis of a recent expansion of the species (last 10,000 years), with colonising populations establishing by migration of a limited number of individuals (founder effect) and/or population dynamics regulated by fire (population bottlenecks). Analysis of the geographic distribution of haplotypic diversity revealed two main groups of Aleppo pine populations: a central Mediterranean group, centred on southern Italy and comprising northern Italian and Spanish populations; and a southern Mediterranean group, centred on Greece. Almost all the haplotypic diversity detected in this species is concentrated in a very limited area located in Greece, which is considered to be the centre of origin of the species from which the recolonization in the post-glacial period started (Morgante *et al.*, 1998). For *P. brutia*, no clear geographic structure was found, even though the degree of genetic differentiation was relatively high (R_{st} of about 30%).

Paternity analysis and detection of natural hybridisation (introgression)

Abies alba

Two relatively isolated adult trees about 30 m apart, as well as 24 naturally regenerated young trees in their direct neighbourhood, were analysed at two chloroplast microsatellite loci (Ziegenhagen *et al.*, 1998). Among all adult and young trees, five different length variants were found for each of the two microsatellites. Observed individual combinations of the size variants of the two loci allowed the definition of five different haplotypes. Figure 9 schematically represents the spatial distribution of the two adult trees (A and B) and the surrounding young trees. Each individual is characterised by the length variants at both microsatellite loci and the relevant two-locus-haplotype. The results reveal that the two adult trees can be distinguished by both microsatellite loci. Comparison of the haplotypes of the adult trees and of the surrounding young trees indicates that, for six out of 16 young trees, paternity of either A or B can unambiguously be excluded. The study demonstrates the potential usefulness of a novel molecular approach towards paternity analysis in a conifer species.

Pinus halepensis and *Pinus brutia*

Three "diagnostic" markers showing size variants clearly distinguishing *P. halepensis* from the *P. brutia* were used to throw light on the occurrence of natural hybridisation in a Turkish sympatric population of *P. halepensis* and *P. brutia* (Bucci *et al.*, 1998). Strong evidence of introgression of "*halepensis*" haplotypes into *P. brutia* seeds (but not *vice versa*) was detected. The overall hybridisation rate was estimated to be about 11 % of the total number of matings analysed (Table 4). Possible explanations for the observed unidirectional, interspecific gene flow are: 1) differential impollination success (for example, in terms of pollen tube growth rate); 2) unbalanced potential pollen donors (for example, due to different stand density and/or differential stage of *P. halepensis* and *P. brutia* individuals); 3) embryo abortion of *P. brutia* (male) x *P. halepensis* (female) hybrids due to a cellular or molecular mechanism of incompatibility. Previously reported evidence on artificial crossings indicates that the last explanation can be considered a useful working hypothesis for further research.

No.	Haplotype	*P. brutia* seeds			*P. halepensis* seeds			Overall	Type
		Embr	Mega	Allo	Embr	Mega	Allo		
1	*107 120 76*	0	2	0	0	0	0	2	*brutia*
2	*108 119 76*	0	6	10	0	0	0	16	*brutia*
3	*108 120 76*	23	25	15	0	0	0	63	*brutia*
4	*108 120 77*	3	0	6	0	0	0	9	*brutia*
5	*108 121 76*	2	1	0	0	0	0	3	*brutia*
6	*108 122 76*	0	1	0	0	0	0	1	*brutia*
7	*109 120 76*	0	1	0	0	0	0	1	*brutia*
8	*112 114 82*	8	0	0	34	30	2	74	*halepensis*
9	*112 115 82*	0	0	0	1	4	37	42	*halepensis*
10	*113 114 82*	0	0	0	1	2	2	5	*halepensis*
Sample size		36	36	72	36	36	127	343	
No. Haplotypes		5	6	12	3	3	15	44	

Table 4: Analysis of the occurrence of natural hybridization in a sympatric populations (Kiranköy - Turkey) of *P. halepensis* and *P. brutia*. Three `diagnostic' cpSSR marker loci (Pt26081, Pt36480, and Pt41093) showing size variants clearly distinguishing *P. halepensis* from *P. brutia* were used. `Embr' and `Mega' refer to embryos and megagametophytes collected from trees of the sympatric population of the two species, `Allo' refers to embryos collected from trees of the allopatric populations of the two species. `Type' refers to the classification of the haplotypes based on the similarity with those detected in allopatric populations. Haplotypes for the same three loci detected in allopatric stands but not in the sympatric population are not reported. Overall sample size and number of haplotypes detected (last two rows) include also the individuals in allopatry not displayed here (from Bucci *et al.*, 1999).

Conclusions

Chloroplast microsatellite analysis represents an extremely useful and informative approach for studying population history, for monitoring gene flow and hybridisation and for identifying areas harbouring high levels of genetic diversity in conifer species. For all the species studied so far, these markers showed a high degree of length polymorphism, both within and among populations. As far as allozyme data exist for comparison and fixation within populations is measured by G_{st}, fixation at cpSSR marker loci is significantly higher than at the allozyme loci. In comparison to species with maternally inherited chloroplast DNA polymorphism, however, G_{st} values for paternally inherited cpSSRs in conifers are lower. Long-range pollen dispersal as well as different mutation rates may be argued for that. This makes paternally inherited highly polymorphic cpSSRs in conifers less suitable for tracing long-range recolonization routes for intraspecific spatial phylogeny. In contrast they are the most advantageous markers for tracing past genetic process like drift or isolation due to the combined property of the high degree of polymorphism and uniparental inheritance, the latter marking one-half the size of reproductively effective population size compared to the parentally inherited markers. Moreover they

are characterised by a high degree of "universality", thus allowing the transfer of primers from one species to others belonging to different taxonomic classifications and therefore obviating the need and expense to develop primers independently for each species. Limited technological investments are required for the analysis of chloroplast microsatellite variation. Moreover, primers and methods developed in one lab can be easily transferred to other laboratories, with a high degree of reproducibility. The approach and the method of detection can also be automated to a large extent, increasing dramatically its efficiency and allowing to obtain a large set of data in a relatively short period of time. The availability of a large set of data represents an important requisite for the development of maps of the distribution of genetic resources of forest tree species: distribution maps are useful tools for an appropriate definition of programmes for the conservation of biodiversity.

References

Bucci G, Anzidei M, Madaghiele A, Vendramin G G (1998) Detection of haplotypic variation and natural hybridization in *halepensis*-complex pine species using chloroplast SSR markers. *Molecular Ecology* 7: 1633-1643.

Bucci G, Vendramin GG (2000) Delineation of genetic zones in the European Norway spruce natural range: preliminary evidences. *Molecular Ecology* (submitted).

Cato SA, Richardson TE (1996) Inter- and intraspecific polymorphism at chloroplast SSR loci and the inheritance of plastids in *Pinus radiata* D. Don. *Theor. Appl. Genet.* 93: 587-592.

Dumolin S, Demesure B, Petit RJ (1995) Inheritance of chloroplast and mitochondrial genomes in pedunculate oak investigated with an efficient PCR method. *Theor. Appl. Genet.* 91: 1253-1256.

Echt CS, De Verno LL, Anzidei M, Vendramin GG (1998) Chloroplast microsatellites reveal population genetic diversity in red pine, *Pinus resinosa* Ait. *Molecular Ecology* 7: 307-317.

Edwards KJ, Barker JHA, Daly A, Jones C, Karp A (1996) Microsatellite libraries enriched for several microsatellite sequences in plants. *BioTechniques* 20: 758-760.

Goldstein DB, Linares AR, Cavalli-Sforza LL, Feldman MW (1995) An evaluation of genetic distances for use with microsatellite loci. *Genetics* 139: 463-471.

Gregorius H-R (1984) A unique genetic distance. *Biometrical Journal* 26: 13-18.

Jarne P, Lagoda JL (1996) Microsatellites, from molecules to population and back. *Tree* 10: 424-429.

Mariette S, Chagne D, Decroocq S, Vendramin GG, Lalanne C, Madur D, Plomion C (2000) Microsatellite markers for *Pinus pinaster* Ait. *Can. J. For. Res.*, submitted.

Morgante M, Felice N, Vendramin GG (1997) Analysis of hypervariable chloroplast microsatellites in *Pinus halepensis* reveals a dramatic genetic bottleneck. In: Karp A, Isaac PG, Ingram DS (eds.). *Molecular Tools for Screening Biodiversity - Plants and Animals.* London: Chapman and Hall, pp. 407-412.

Nei M (1973) Analysis of gene diversity in subdivided populations. *Proc. Natl. Acad. Sci. USA* 70: 3321-3323.

Powell W, Morgante M, McDevitt R, Vendramin GG, Rafalski JA (1995) Polymorphic simple sequence repeat regions in chloroplast genomes: Applications to the population genetics of pines. *Proc. Natl. Acad. Sci. USA* 99: 7759-7763.

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139: 457-462.

Soranzo N, Provan J, Powell W (1999) An example of microsatellite length variation in the mitochondrial genome of conifers. *Genome* 42: 158-161.

Sperisen C, Büchler U, Mátyás G, Anzidei M, Madaghiele A, Skrøppa T, Vendramin GG (1998) Polymorphic tandem repeats in the chloroplast and mitochondrial genomes of Norway spruce. In: Skrøppa T, Paule L, Gömöry D (eds.). *Genetics and Breeding of Norway Spruce.* Arbora Publisher, Zvolen, Slovakia, pp. 15-24.

Tautz D (1993) Notes on the definition and nomenclature of tandemly repetitive DNA sequences. In: Pena SDJ, Chakraborty R, Epplen JT, Jeffreys AJ (eds.). *DNA Fingerprinting: State of the Science.* Birkhauser Verlag, Basel, Switzerland, pp. 21-28.

Valdes AM, Slatkin M, Freiner NB (1993) Allele frequencies at microsatellite loci: the stepwise mutation model revisited. *Genetics* 133: 737-749.

Vendramin GG, Lelli L, Rossi P, Morgante M (1996) A set of primers for the amplification of 20 chloroplast microsatellites in *Pinaceae*. *Molecular Ecology* 5: 111-114.

Vendramin GG, Ziegenhagen B (1997) Characterisation and inheritance of polymorphic plastid microsatellites in *Abies*. *Genome* 40: 857-864.

Vendramin GG, Degen B, Petit RJ, Anzidei M, Madaghiele A, Ziegenhagen B (1999) High level of variation at *Abies alba* chloroplast microsatellite loci in Europe. *Molecular Ecology* 8: 1117-1126.

Vendramin GG, Anzidei M, Madaghiele A, Sperisen C, Bucci G (2000) Chloroplast microsatellite analysis reveals the presence of population subdivision in Norway spruce (*Picea abies* K.). *Genome*, in press.

Wakasugi T, Tsudzuki J, Ito S, Shibata M, Sugiura M (1994) A physical map and clone bank of the black pine (*Pinus thunbergii*) chloroplast genome. *Plant Mol. Biol. Rep.* 12: 227-241.

Weising K, Gardner C (1999) A set of conserved PCR primers for the analysis of simple sequence repeat polymorphism in chloroplast genomes of dicotyledonous angiosperms. *Genome* 42: 9-19.

Ziegenhagen B, Scholz F, Madaghiele A, Vendramin GG (1998) Chloroplast microsatellites as markers for paternity analysis in *Abies alba*. *Can. J. For. Res.* 28: 317-321.

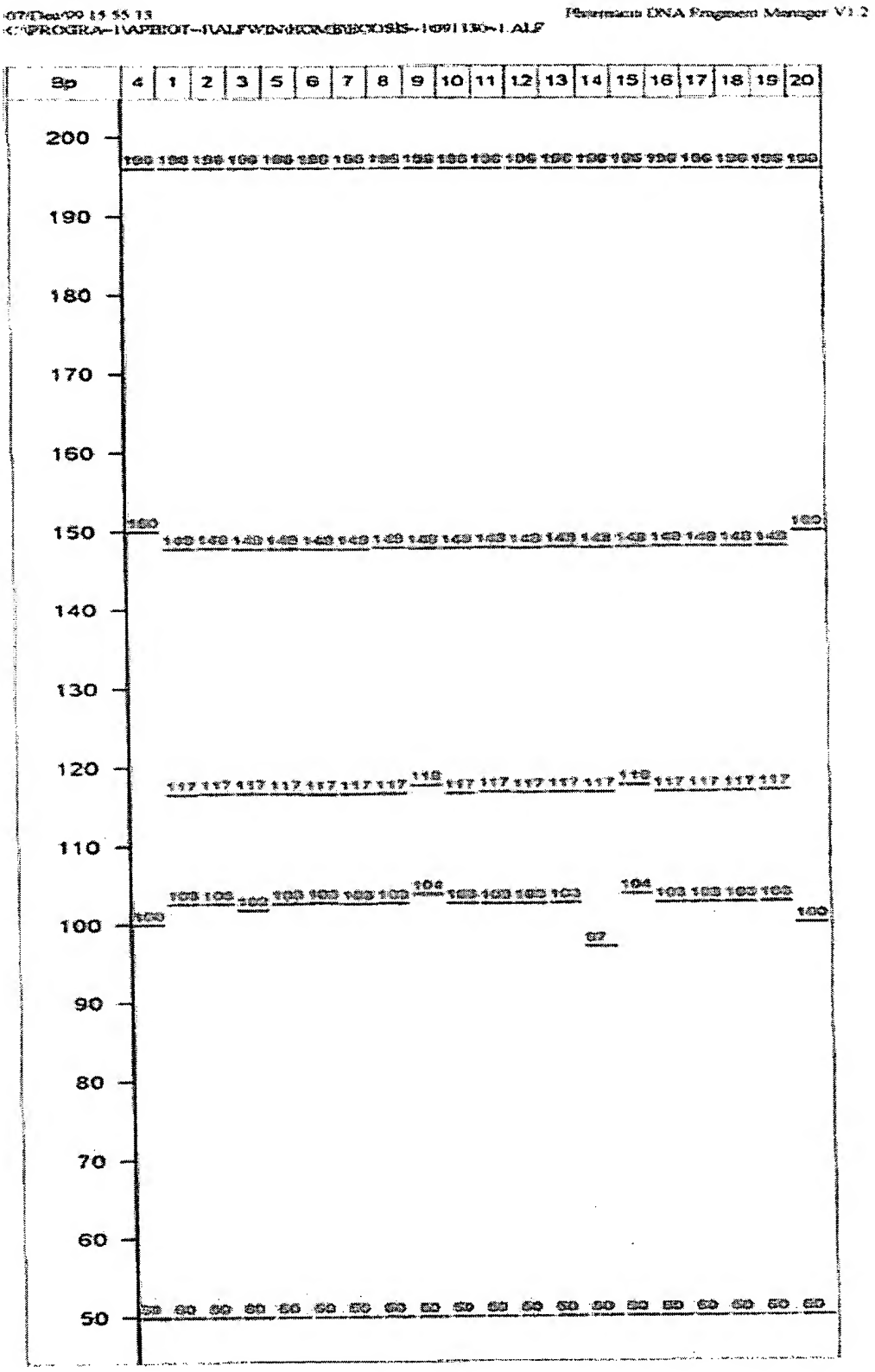


Figure 1: Chloroplast SSR fragments displayed of the Fragment Manager software on a Pharmacia A.L.F. automated sequencer. PCR reactions were performed as described in the text. The Fragment manager vs. 1.2 conversion software was used to detect and size the peaks of fluorescence. Lanes 4 and 20 contain a ladder that is used as external standard. Lanes 1, 2, 3, 5-19: amplifications of three chloroplast microsatellites of *Picea abies* individuals. Fragment sizes are displayed above the

corresponding band. The 50- and 196-bp constant bands in all the lanes are fragments of known size used as internal standards and loaded together with the samples.

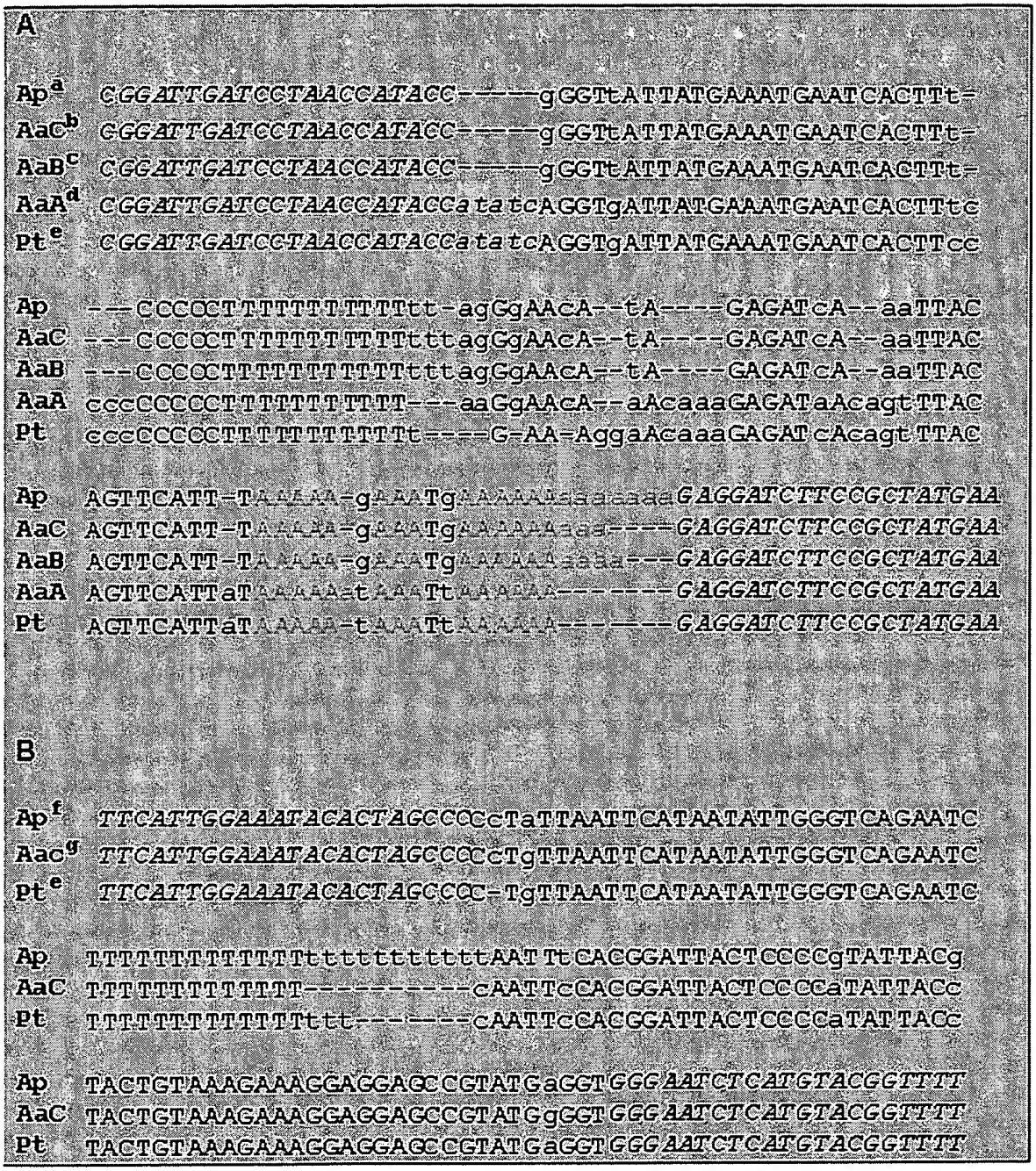


Figure 2: Alignment of cpDNA sequences. **A** Locus Pt 30204, **B** Locus Pt 71936.

Ap (*Abies pinsapo*), AaA (*Abies alba* A), AaB (*Abies alba* B), AaC (*Abies alba* C), and Pt (*Pinus thunbergii*). Types in italics are sequences of the primers, lowercase types indicate nucleotide substitutions and dashes stand for deletions. Coloured sequences indicate the microsatellite stretches.

[a] Vendramin and Ziegenhagen, unpublished, [b] Vendramin and Ziegenhagen, unpublished, [c] GenBank accession No. U82922, [d] Vendramin and Ziegenhagen, unpublished [e] DDBV accession No. D11467, [f]

GenBank accession No. U2923, [g] GenBank accession No. U2924 (from Vendramin and Ziegenhagen, 1997).

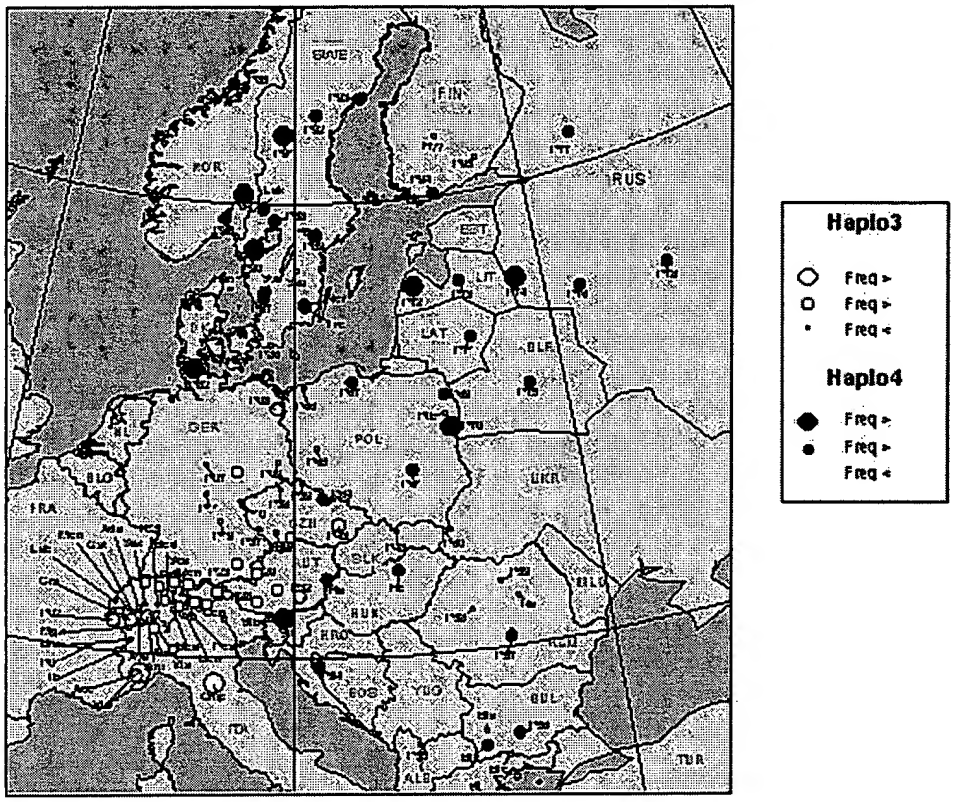


Figure 3: Geographic distribution of cpSSR haplotypes *116/96/144* (Haplo03) and *116/100/143* (Haplo04, allele size at the three loci Pt 26081, Pt 63718, and Pt 71936, respectively) across the Norway spruce natural range. Circles representing the populations are proportional to the relative within-stand haplotype frequency (from Vendramin *et al.*, 2000).

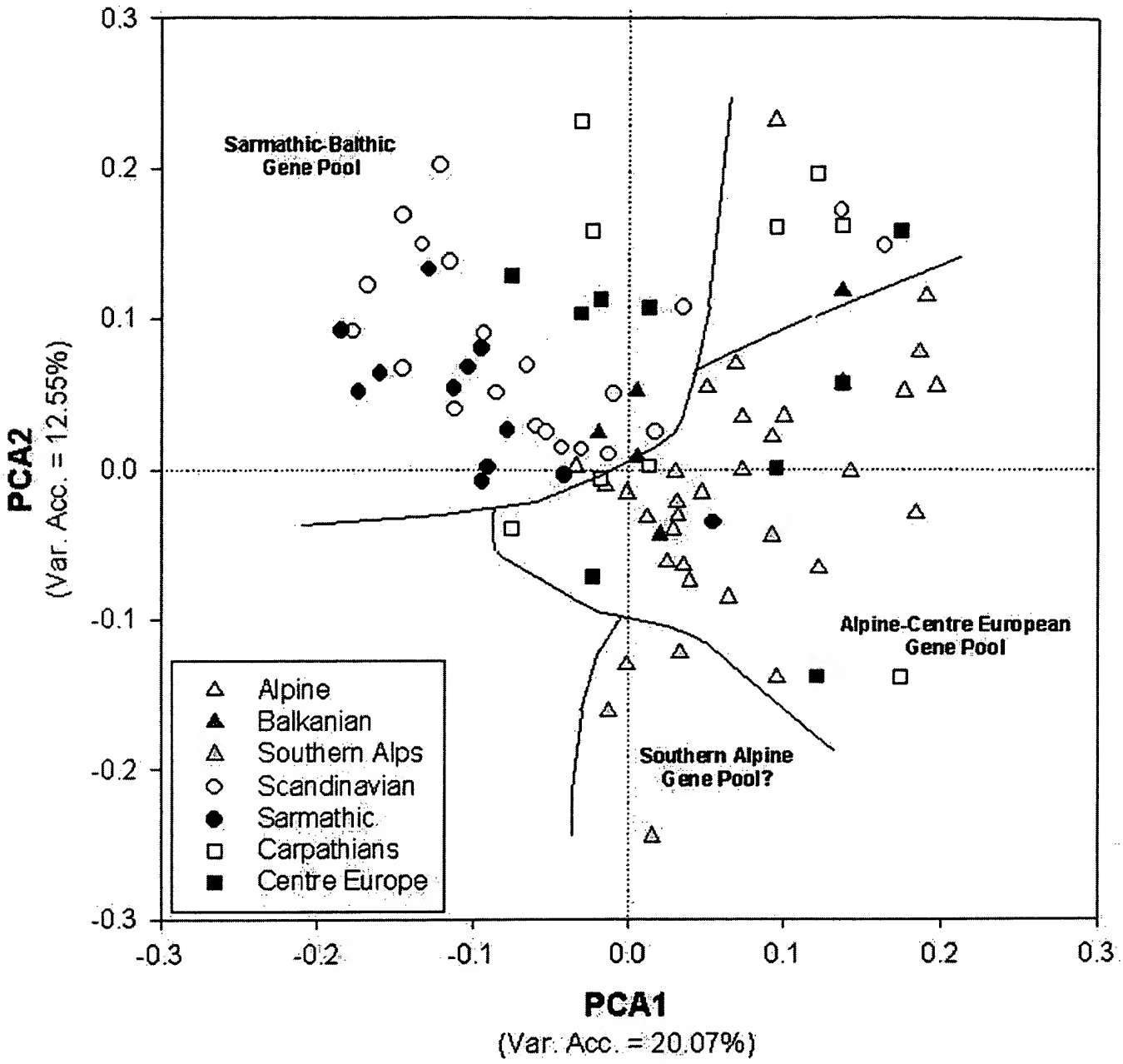


Figure 4: Plot of the first two components of the standardized PCA carried out on transformed haplotype frequencies. Stands belonging to the same geographic area are shown by the same symbol. Despite the large genetic noise and the continuous variation across the range, a fairly good separation between two main group of populations ("Sarmathic-Baltic" group, including population from western Russia and Fennoscandia, and "Alpine - Centre European" group, including populations from Italy, Switzerland, Austria, Slovenia) can be recognized. Balkanian stands were clustered at the boundary between the two main gene pools, while centre-European and Carpathian populations were scattered all over the two groups. A third, possible group of populations including stands from south-western Alps was detected. Lines delimiting gene pools were drawn arbitrarily (from Vendramin *et al.*, 2000).

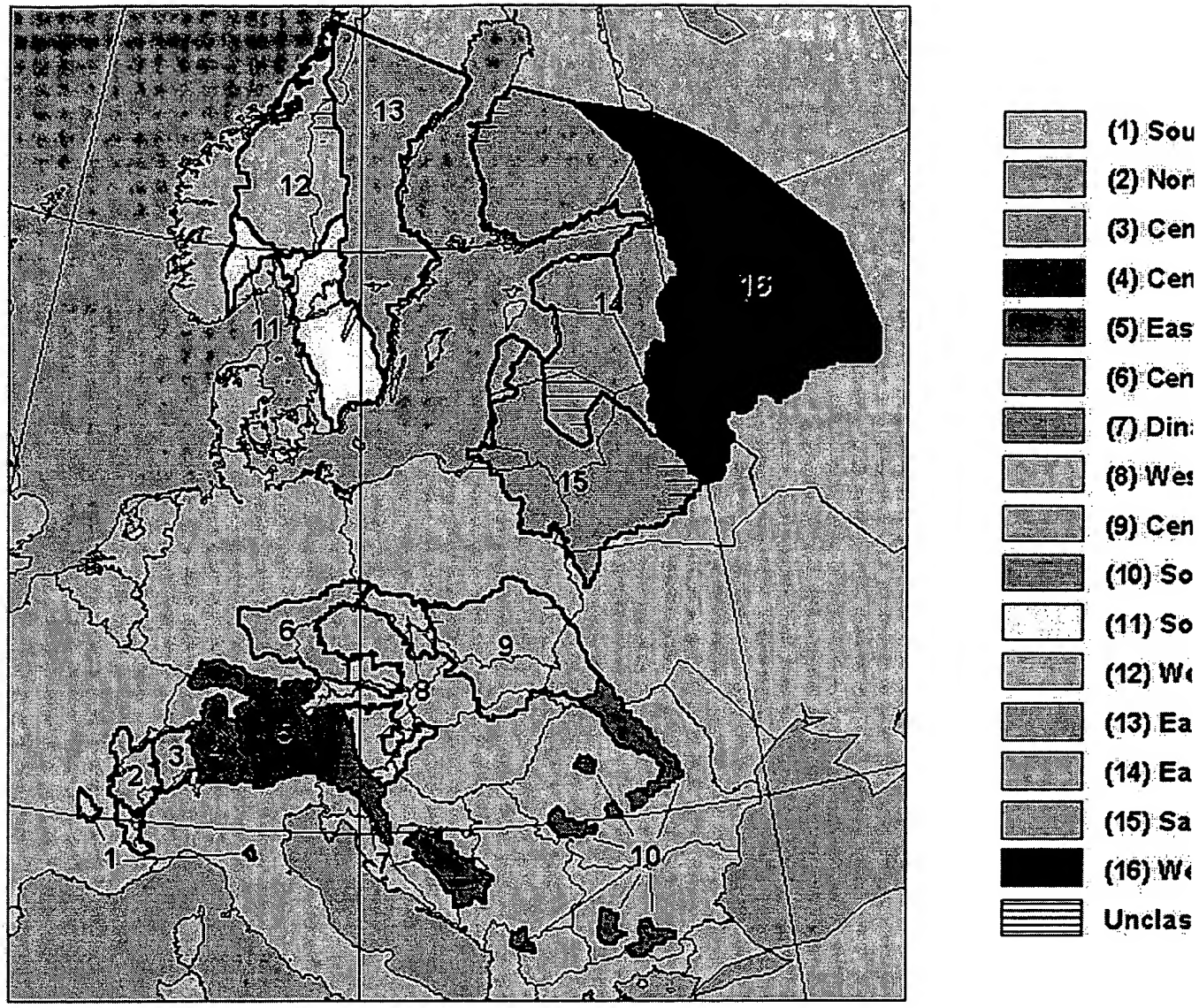


Figure 5: Results of the k-means cluster analysis carried out on interpolated haplotype-frequency surfaces (from Bucci and Vendramin, 2000).

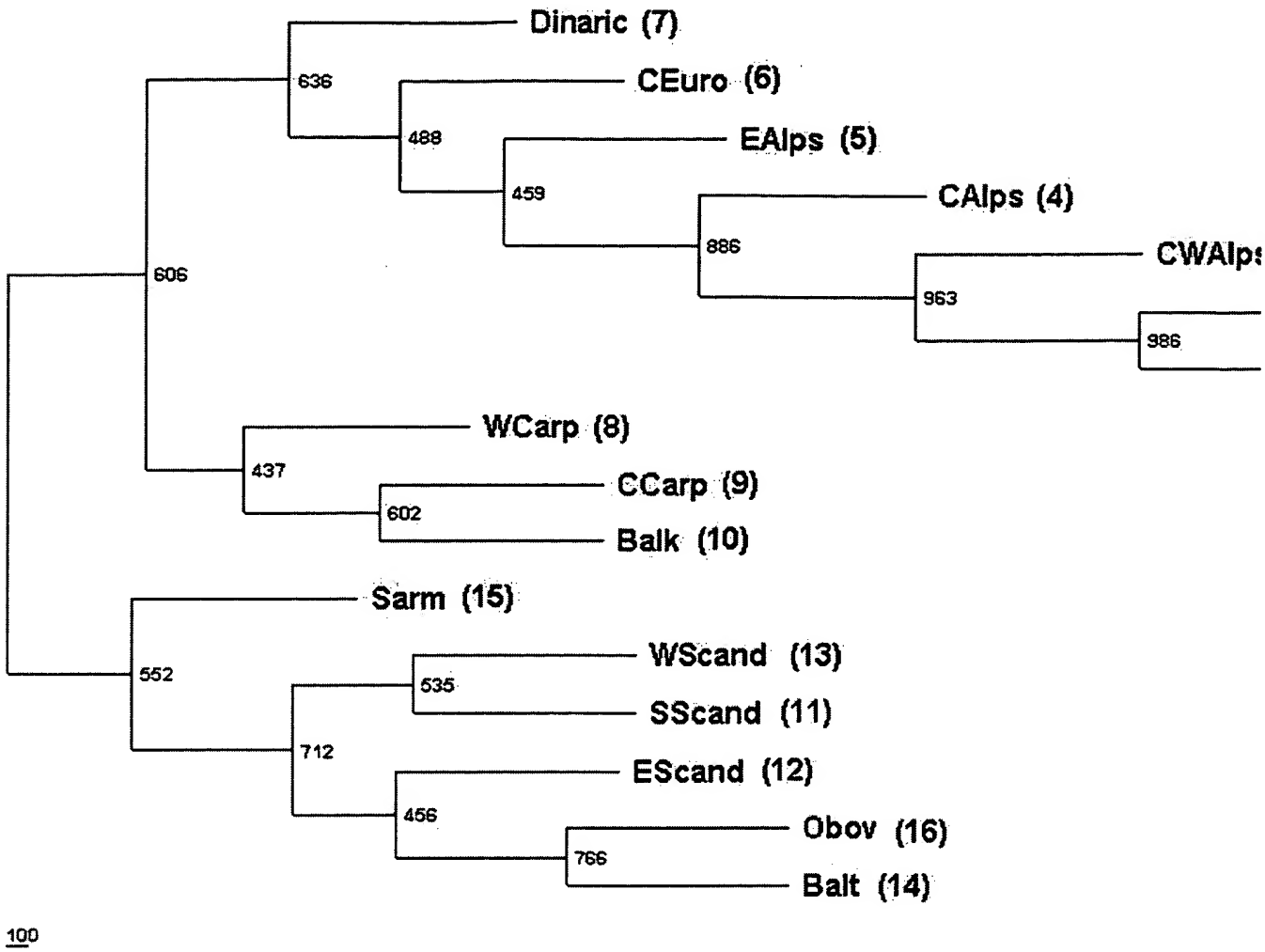


Figure 6: Majority-rule consensus tree obtained by restricted maximum likelihood method (reml) on transformed mean haplotype frequencies for each genetic zone. Bootstrap values (N = 1000) are indicated at each node (from Bucci and Vendramin, 2000).

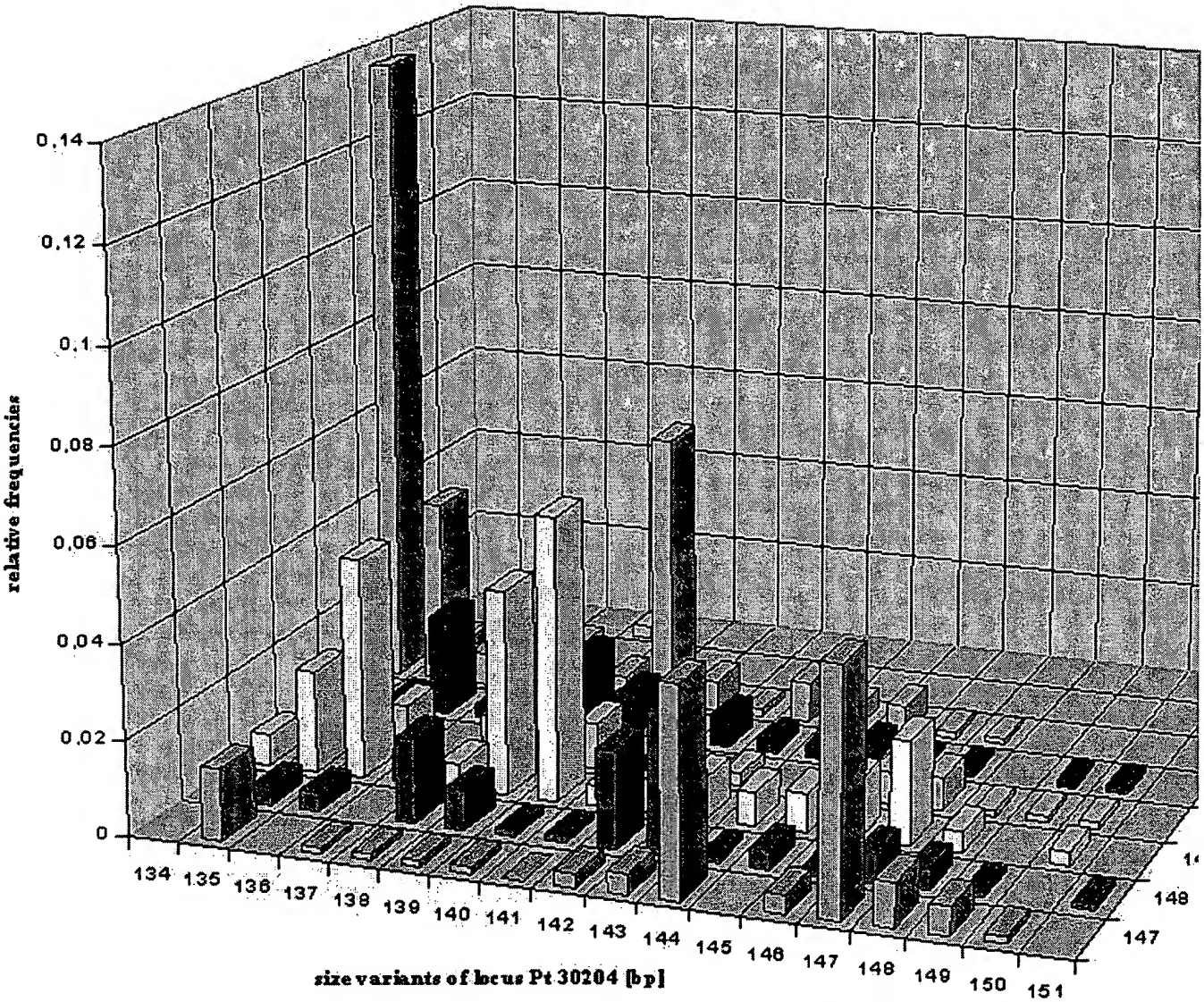
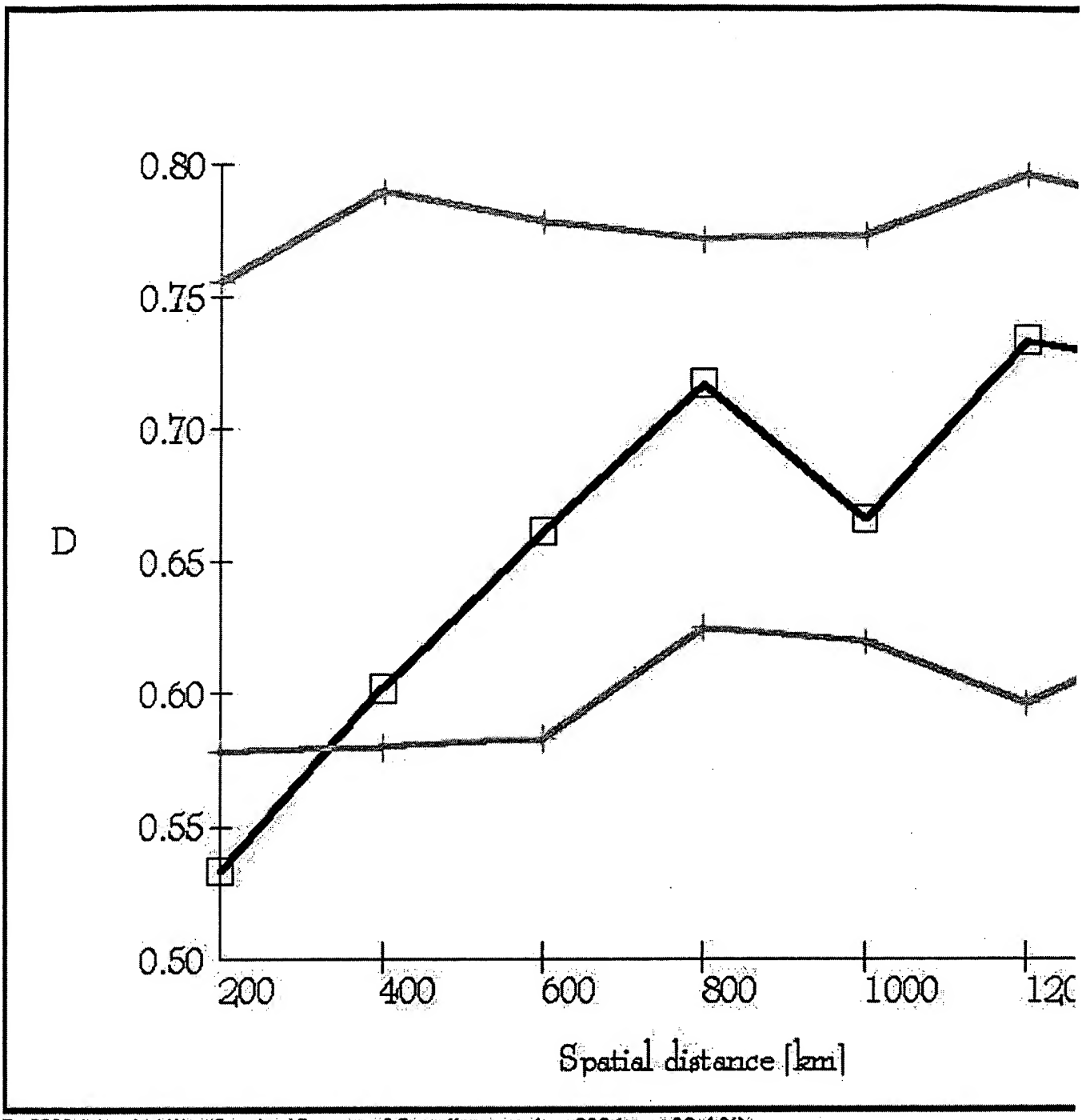
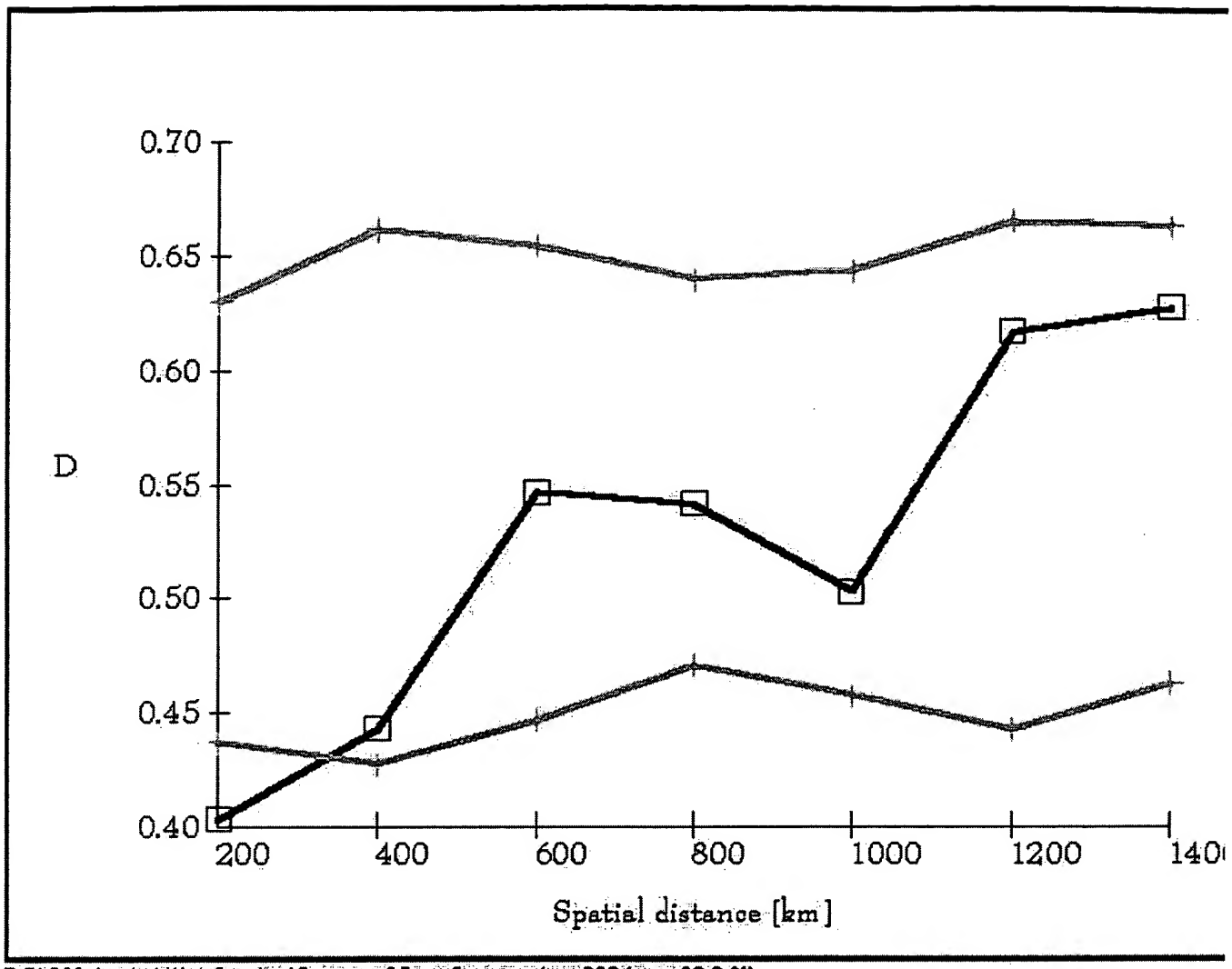


Figure 7: Relative frequencies of two-locus-combinations (haplotypes) identified in the 17 investigated silver fir populations (from Vendramin *et al.*, 1999).



Pt 30204 (probability for significance of first distance class 200 km = 99.4 %)

(8a)



Pt71936 (probability for significance of first distance class 200 km = 98.2 %)

(8b)

Figure 8: Distogram of the average (D) of the genetic distances d_0 (Gregorius 1984) of all pairs of individuals belonging to each of seven spatial distance classes. The 90% confidence interval of 1000 permutations is presented. **(a)**: Distogram of average genetic distance D estimated for locus Pt 30204. **(b)**: Distogram of average genetic distance D estimated at locus Pt 71936 (from Vendramin *et al.*, 1999).

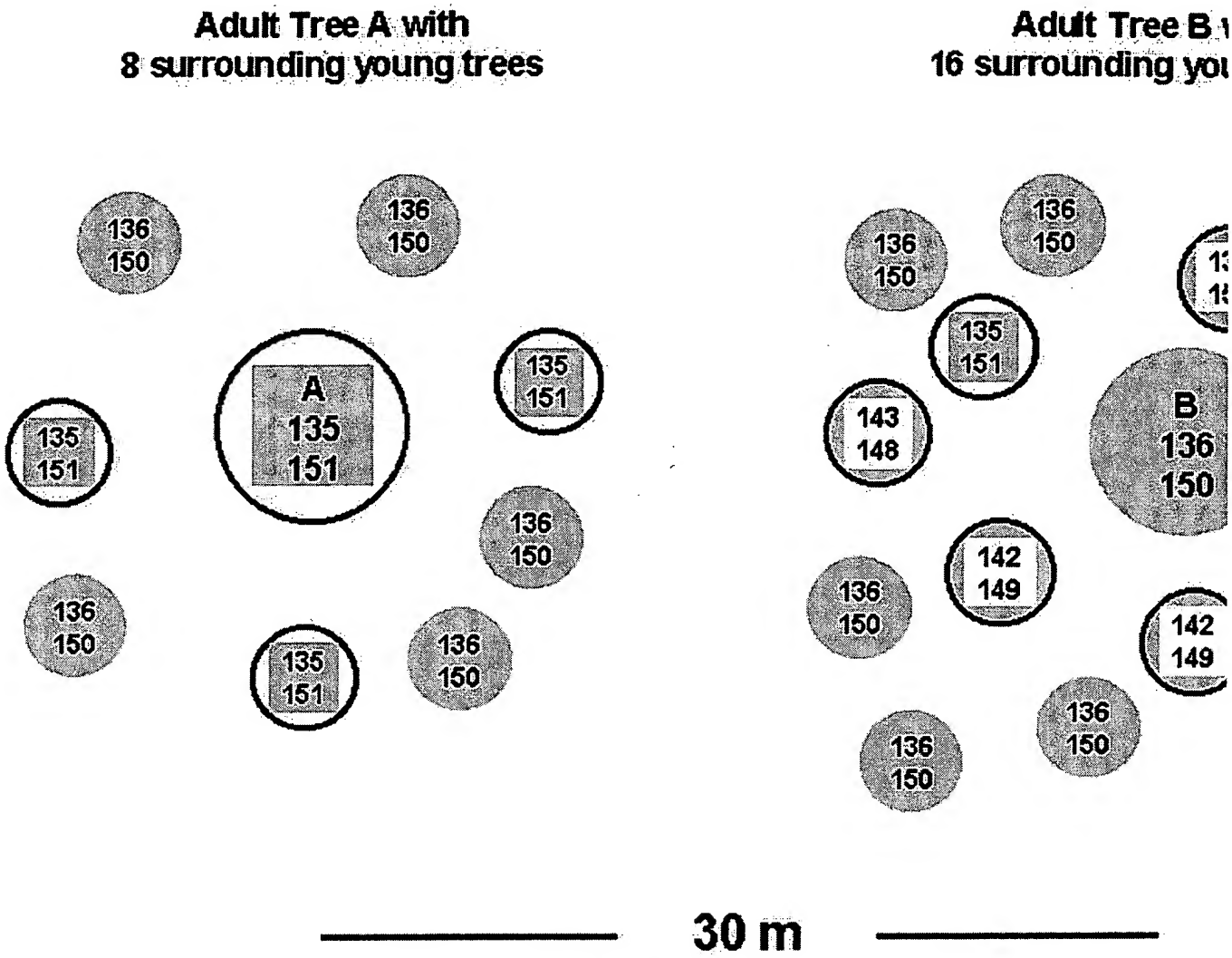


Figure 9: Schematical drawing of two isolated adult *Abies alba* trees A and B with surrounding young *Abies alba* trees. Circles represent individual trees, numbers in circles give the size variants in basepairs at two chloroplast microsatellite loci Pt 30204 and Pt 71936 (Vendramin and Ziegenhagen 1997b). Length variants of Pt 30204 are located at the top, length variants of Pt 71936 at the bottom of each circle, the combination yielding the relevant haplotype. Three different symbols (types of shading) were chosen: two for the individuals carrying the two different haplotypes of A and B, and one for characterizing individuals with all other haplotypes different from A and B (from Ziegenhagen *et al.*, 1998).

Plant Growth Regulation 20: 47–52, 1996.

Short paper

Application of AFLP™-based molecular markers to breeding of *Populus* spp.

María Teresa Cervera[1], Jaqueline Gusmão[1], Marijke Steenackers[2], Ann Van Gysel[1], Marc Van Montagu[1]* & Wout Boerjan[1]

[1]*Laboratorium voor Genetica, Department of Genetics, Flanders Interuniversity Institute for Biotechnology, Universiteit Gent, K.L. Ledeganckstraat 35, B-9000 Gent, Belgium;* [2]*Instituut voor Bosbouw en Wildbeheer, B-9500 Geraardsbergen, Belgium (*author for correspondence, tel: 32-9-2645170; fax: 32-9-2645349; email: mamon@gengenp.rug.ac.be)*

Key words: AFLP™, fingerprint, genome mapping, molecular markers, poplar, QTL

Abstract

Molecular marker technologies have eased and potentiated genetic analysis of plants and have become an extremely useful tool in forest tree breeding. The information provided by molecular markers has made it possible to acquire further knowledge about the structure and organization of plant genomes as well as about the evolution of these plant genomes through phylogenetic analysis. Using *Populus* spp. as a model tree, this paper aims at showing and discussing the possible applications of AFLP™, a high-density DNA marker technology developed by Keygene N.V. (Wageningen, The Netherlands). Applications include: (i) AFLP analysis of the disease resistance against *Melampsora larici-populina* using bulked-segregant analysis, (ii) AFLP fingerprinting for identification and taxonomic analysis of individual trees, and (iii) AFLP-based mapping strategies in *Populus*.

Abbreviations: AFLP = amplified fragment length polymorphism; RFLP = restriction fragment length polymorphism; PCR = polymerase chain reaction; QTL = quantitative trait loci; RAPD = random amplified polymorphic DNA

Introduction

The genus *Populus* shows wide genetic variation. *Populus* spp. are native to the Northern hemisphere (North America, Europe, North Africa, and Asia). Thirty species of poplars, cottonwoods, and aspens have been described. Interspecific as well as intraspecific hybrids are easily obtained and often display an extensive heterosis effect.

Poplar is considered a model tree in all aspects of forest tree biology. In general, *Populus* trees are easy to grow and can be vegetatively propagated from stem or root cuttings, thereby enabling a rapid multiplication for experimental or commercial use. Due to its rapid growth, poplar (*Populus* spp.) has become a tree of high economic importance. A wide information about the genetic material is available, multi-generation pedigrees exist, and large-scale screenings of parents and respective progenies segregating for traits of scientific or commercial interest have been performed. In Europe, poplar wood is used primarily for the construction of boxes, pallets, soft board, and multiplex, whereas in the United States and Canada poplar wood is mainly processed to pulp and paper.

Due to the advantageous molecular characteristics, genetic manipulations are more accessible in poplar than in most forest trees; the nuclear genome is relatively small (2C = 1.1 pg) and the chromosome number of all species is identical (2n = 38) [5, 42]. Several species of poplar can be transformed either by *Agrobacterium tumefaciens* infection or by direct DNA transfer procedures [17]. A fast and efficient method to

transform *P. tremula* × *P. alba* by *Agrobacterium* has been described [22].

Different molecular techniques, such as restriction fragment length polymorphism (RFLP) and polymerase chain reaction (PCR)-based techniques, have been applied with the aim of detecting molecular markers associated with qualitative or quantitative trait loci (QTLs) of poplar, especially those related to disease resistance and economically important traits [5, 7]. This approach should result in powerful and fast methods for indirect selection.

The identification of molecular markers associated to a particular trait is crucial for the breeder because the selection for this trait can then be performed at very early stages of development. Moreover, the amount of progeny plants to be scored can be increased dramatically. This should allow the isolation of individuals carrying a range of common traits of interest yet genetically diverse for the rest of their genome. In this way, multiclonal plantations can be obtained where the individuals have the same characteristics but remain polymorphic. A high biodiversity in the plantation is expected to increase the scope of resistance against new pathogens. In addition, molecular markers will assist breeders in the choice of parents for new breeding programs.

Molecular markers facilitate the introgression of traits. The aim of different breeding programs is to introgress a gene, or genes, of interest from a donor parent into an elite parent, generally through back-crossing. Molecular markers are used to monitor the presence or absence of the locus to introgress in a segregating population and also facilitate the selection of those back-crossings that are genetically the most similar to the recurrent parent [14, 25].

High levels of heterozygosity shown by individual trees together with the availability of F_1 families and a molecular marker technology are sufficient to generate genomic maps for forest tree species. Genomic maps provide information about genome structure, organization and evolution. Among related species, comparative analysis of genetic linkage maps (syntenic mapping) allows the determination, by using heterologous probes, of the relative order of genes along the chromosomes [1, 2, 4, 19, 20, 21, 38, 40, 43, 44]. The generation of high-density molecular maps permits the identification of molecular markers tightly linked to a locus of interest, which may lead to map-based cloning of the gene(s) of interest by chromosome landing [39].

Recently, several high-density marker technologies have been developed. Among them, AFLP™ [1] is considered one of the most powerful technologies. This method was developed at Keygene N.V. (Wageningen, The Netherlands) by Zabeau and Vos [41, 45]. AFLP markers assay the presence/absence of restriction enzyme sites and sequence polymorphisms adjacent to these sites. Briefly, three crucial steps are followed to obtain AFLP markers: (i) digestion of genomic DNA with two different enzymes, such as *Mse*I (frequent-cutter enzyme) and *Eco*RI (rare-cutter enzyme); (ii) ligation of adapter oligonucleotides to the restriction ends; and (iii) selection of fragments by two successive PCR-based amplification steps using primers complementary to the adapter oligonucleotides with additionally one to three selective nucleotides.

AFLP has several advantages over random amplified of polymorphic DNA (RAPD). First, a higher number of loci can be analyzed per experiment; approximately 10-fold the number of informative markers are obtained per analysis. Second, AFLP markers are co-dominant; by using the appropriate equipment and software to analyze the gels, it is possible to identify whether an allele is homo- or heterozygous. This provides more information than dominant markers such as RAPDs [41]. Third, AFLP gives highly reproducible banding patterns due to a highly specific annealing of the primers to the complementary adapter oligonucleotides [41]. The RAPD technology, on the other hand, can suffer from a lack of reproducibility which is caused by mismatch annealing of the random primers.

AFLP analysis of disease resistance in *Populus*

Disease resistance has always been one of the most important selection criteria in poplar breeding. Clones selected by the breeders are vegetatively propagated to generate monoclonal forests which, unfortunately, are known to be fragile with regard to new pathogen attacks [30].

The most important diseases damaging poplar species in Central and Northern Europe are leaf rust caused by the fungus *Melampsora larici-populina*, bacterial canker, caused by *Xanthomonas populi*, and leaf spot, caused by the fungus *Marssonina brunnea*. The infection produced by *M. larici-populina* causes premature defoliation and can reduce growth by more

[1] AFLP is a registered trademark in the Benelux.

than 20%. Trees defoliated early in the growing season become more susceptible to secondary pathogen infections and to environmental stress [30]. Repeated infections during successive years can result in a complete loss of the plantation.

We have identified molecular markers tightly linked to the locus conferring resistance against *M. larici-populina* in *Populus* applying the AFLP technique [10]. Segregation of the resistance against *M. larici-populina* in the F_1 progeny (family 87001) of a controlled cross between a resistant female, *P. deltoides* (clone S 9–2), and a susceptible male, *P. nigra* (Ghoy), is consistent with the combination of two different factors: a monogenic resistance caused by a single dominant resistance gene, resulting in 50% resistant progeny, and a multigenic (horizontal) resistance displayed by different degrees of susceptibility shown by the other 50%. Our efforts have focused on the identification of molecular markers co-segregating with the monogenic resistance. To facilitate the screening of family 87001 and to increase the chance of identifying molecular markers that are closely linked to the resistance locus, we decided to used the bulked segregant analysis (BSA) [27]. This method is based on bulking DNAs from segregating populations. For our analysis, two bulks of DNA were prepared by pooling equal amounts of DNA extracted from resistant plants ("resistant bulk") and from susceptible plants ("susceptible bulk"). Using 144 primer combinations, we identified three different AFLP markers present in the resistant parent and the resistant bulk, but absent in the susceptible parent and the susceptible bulk. Linkage of these markers to the resistance gene was confirmed by AFLP analysis of each individual DNA from the family 87001 (Figure 1) [10].

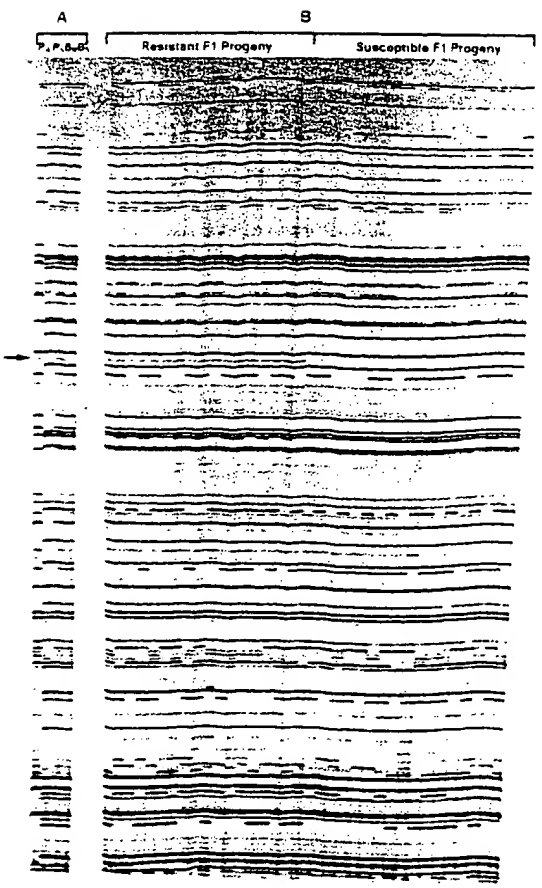


Figure 1. Identification of an AFLP marker linked to a *Melampsora* resistance locus in *Populus*. (A) The bulked segregant analysis (BSA) is presented as a set of four lanes: resistant parent (P_R), susceptible parent (P_S), resistant bulk (B_R), and susceptible bulk (B_S). (B) The AFLP marker, identified by BSA and indicated by an arrow, is present in resistant but absent in susceptible F_1 progeny (family 87001).

AFLP fingerprinting as a tool for identification and taxonomic analysis

Obtaining genomic fingerprints for each species of *Populus* has become a powerful tool to discriminate individual genotypes and to determine phylogenetic relationships among these species. Traditional taxonomy based on morphological characters [15], isozyme analysis [23, 32, 33, 37] and gas chromatography [13] revealed the first estimations of genetic and taxonomic relationships between *Populus* species but was often strongly influenced by the environment. RFLP technology has contributed to the identification of DNA polymorphisms as genetic markers in *Populus*. In this way RFLP genome analysis [8, 18, 23], mitochondrial DNA analysis [3, 31], chloroplast DNA studies [26, 34–36], and ribosomal DNA analysis [11] have been performed in different *Populus* species. Recently, PCR-based marker systems, such as sequence-tagged site (STS), RAPD, or simple sequence repeat (SSR) have caused a revolution in genome analysis; two of these methods, RAPD and STS have been used in genetic studies of *Populus* spp. [7–9, 23].

Given the high level of genetic polymorphism between *Populus* species and the advantages offered by AFLP analysis (see Introduction), studies on diversity between *Populus* spp. using AFLP will probably result in a highly reliable classification. Such an analysis is

being carried out in our laboratory using plant material provided by the IBW-Geraardsbergen (former Institute for Poplar Culture) in Belgium, which has a large collection of species originating from Asia, North America, and Europe. Using AFLP, approximately 60 to 100 markers (bands per lane) can be obtained per primer combination and, on average, 60% of these bands are polymorphic between different species (*P. deltoides* and *P. nigra*) and 15–25% between two random full-siblings derived from an interspecific cross *P. deltoides* × *P. nigra*. These data indicate that it should be possible to establish reliable taxonomic relationships of different *Populus* spp. as well as different genotypes of the same species.

Using AFLP, patterns characteristic for a specific cross, including the fingerprint of an individual tree, can be obtained thereby reducing expensive and difficult storage of nearly identical material. Well managed, this can also be used as a tool to increase population diversity. Additionally, fingerprinting of individual trees will help the breeder to legally protect new genotypes that are to be released on the market.

Genome mapping in *Populus*

Genome maps allow the localization of loci controlling monogenic or multigenic traits. Complex trait mapping provides information about the number, chromosomal location, magnitude of effect and interactions of genetic loci controlling the expression of a particular trait. In the near future, it should be possible to clone, by chromosome landing, genes controlling characters that are quantitatively inherited [16, 29].

The first linkage groups identified in *Populus* were obtained using alloenzymes [24, 28] and RFLPs [24]. Recently, a new *Populus* linkage map has been reported [8]. This map contains 343 DNA-based markers, combining RFLPs, STSs, and RAPDs, with which an F_2 progeny of interspecific hybrids was mapped. QTLs with large effects on stem growth and form, two important commercial traits, and on spring leaf phenology, an adaptive trait, have been mapped [6–8].

We are constructing linkage maps of single individual *Populus* trees using the "two-way pseudo-testcross" strategy [12] in combination with the AFLP technology. The "two-way pseudo-testcross" mapping strategy is based on linkage analysis of those markers that segregate 1:1 in an F_1 full-sib family, i.e. those markers that are heterozygous in one parent and null in the other parent. In this way, two genome maps are constructed, one for each parent. Given the high level of heterozygosity in individual *Populus* trees and the high number of markers generated by the AFLP technique, we expect to obtain high-density linkage maps of *P. deltoides*, *P. nigra* and *P. trichocarpa*. These genome maps will be obtained by analyzing two different F_1 families (87001 and 87002) derived from controlled interspecific crosses and sharing a common female parent (*P. deltoides*). They will allow the genetic mapping of resistance to pathogens such as *Melampsora* spp., *Xanthomonas populi*, poplar mosaic virus and *Marssonina brunnea*, as well as other economically important traits such as those related to stem growth, leaf phenology, shape, wood density and frost damage.

Acknowledgments

Experimental work has been carried out according to protocols from Keygene N.V. The authors would like to thank Vic Steenackers and Jos Van Slycken (IBW-Geraardsbergen) and Marc Zabeau (Keygene N.V., Wageningen, The Netherlands) for fruitful long-standing collaborations, Carlos A. Malpica for his valuable comments, Ann Vanden Broeck (IBW-Geraardsbergen) for providing part of the plant material described, Martine De Cock for help with the manuscript, and Karel Spruyt for photographic work. This work was supported by grants from the Flemish Government (BNO/BB/6/1994) and the Commission of the European Communities AIR program (AIR1-CT92-0349), called Inter-disciplinary Research for Poplar Improvement (IRPI). M.-T.C. is funded by an individual fellowship from the Human Capital and Mobility program of the European Union (41AS8694).

References

1. Ahn S, Anderson JA, Sorrells ME and Tanksley SD (1993) Homoeologous relationships of rice, wheat and maize chromosomes. Mol Gen Genet 241: 483–490
2. Ahn S and Tanksley SD (1993) Comparative linkage maps of the rice and maize genomes. Proc Natl Acad Sci USA 90: 7980–7984
3. Barrett JW, Rajora OP, Yeh FCH, Dancik BP and Strobeck, C (1993) Mitochondrial DNA variation and genetic relationships of *Populus* species. Genome 36: 87–93
4. Bonierbale MW, Plaisted RL and Tanksley, SD (1988) RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. Genetics 120: 1095–1103

5. Bradshaw HD Jr and Stettler RF (1993) Molecular genetics of growth and development in *Populus*. I. Triploidy in hybrid poplars. Theor Appl Genet 86: 301–307
6. Bradshaw HD Jr and Stettler RF (1994) Molecular genetics of growth and development in *Populus*. II. Segregation distortion due to genetic load. Theor Appl Genet 89: 551–558
7. Bradshaw HD Jr and Stettler RF (1995) Molecular genetics of growth and development in *Populus*. IV. Mapping QTLs with large effects on growth, form and phenology traits in a forest tree. Genetics 139: 963–973
8. Bradshaw HD Jr, Villar M, Watson BD, Otto KG, Stewart S and Stettler RF (1994) Molecular genetics of growth and development in *Populus*. III. A genetic linkage map of a hybrid poplar composed of RFLP, STS, and RAPD markers. Theor Appl Genet 89: 167–178
9. Castiglione S, Wang G, Damiani G, Bandi C, Bisoffi S and Sala, F (1993) RAPD fingerprints for identification and for taxonomic studies of elite poplar (*Populus* spp.) clones. Theor Appl Genet 87: 54–59
10. Cervera M-T, Gusmão J, Steenackers M, Peleman J, Storme V, Vanden Broeck A, Van Montagu M, Boerjan W (1996) Identification of AFLP molecular markers for resistance against *Melampsora larici-populina* in *Populus*. Theor Appl Genet: in press
11. Faivre-Rampant P, Jeandroz S, Lefevre F, Lemoine M, Villar M and Berville A (1992) Ribosomal DNA studies in poplars: *Populus deltoides*, *P. nigra*, *P. trichocarpa*, *P. maximowiczii*, and *P. alba*. Genome 35: 733–740
12. Grattapaglia D and Sederoff R (1994) Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-testcross: mapping strategy and RAPD markers. Genetics 137: 1121–1137
13. Greenway W, Jobling J and Scaysbrook T (1989) Composition of bud exudates of Populus × interamericana clones as a guide to clonal identification. Silvae Genetica 38: 29–32
14. Hillel J, Schaap T, Haberfeld A, Jeffreys AJ, Plotzky Y, Cahaner A and Lavi U (1990) DNA fingerprints applied to gene introgression in breeding programs. Genetics 124: 783–789
15. Hu CC, Crovello TJ and Sokal RR (1985) The numerical taxonomy of some species of Populus based only on vegetative characters. Taxon 34: 197–206
16. Jansen RC and Stam P (1994) High resolution of quantitative traits into multiple loci via interval mapping. Genetics 136: 1447–1455
17. Jouanin L, Brasileiro ACM, Leplé JC, Pilate G and Cornu D (1993) Genetic transformation: a short review of methods and their applications, results and perspectives for forest trees. Ann Sci For 50: 325–336
18. Keim P, Paige KN, Whitham TG and Lark KG (1989) Genetic analysis of an interspecific hybrid swarm of Populus: occurrence of unidirectional introgression. Genetics 123: 557–565
19. Kowalski SP, Lan T-H, Feldmann KA and Paterson AH (1994) Comparative mapping of *Arabidopsis thaliana* and *Brassica oleracea* chromosomes reveals islands of conserved organization. Genetics 138: 499–510
20. Kurata N, Moore G, Nagamura Y, Foote T, Yano M, Minobe Y and Gale M (1994) Conservation of genome structure between rice and wheat. Bio/technology 12: 276–278
21. Lagercrantz U, Putterill J, Coupland G, and Lydiate D (1996) Comparative mapping in *Arabidopsis* and *Brassica*, fine scale genome collinearity and congruence of genes controlling flowering time. Plant J 9: 13–20
22. Leplé JC, Brasileiro ACM, Michel MF, Delmotte F and Jouanin L (1992) Transgenic poplars: expression of chimeric genes using four different constructs. Plant Cell Rep 11: 137–141
23. Liu Z and Furnier GR (1993) Comparison of allozyme, RFLP, and RAPD markers for revealing genetic variation within and between trembling aspen and bigtooth aspen. Theor Appl Genet 87: 97–105
24. Liu Z and Furnier GR (1993) Inheritance and linkage of allozymes and restriction fragment length polymorphisms in trembling aspen. J Hered 84: 419–424
25. Mazur BJ and Tingey SV (1995) Genetic mapping and introgression of genes of agronomic importance. Curr Opin Biotechnol 6: 175–182
26. Mejnartowicz M (1991) Inheritance of chloroplast DNA in *Populus*. Theor Appl Genet 82: 477–480
27. Michelmore RW, Paran I and Kesseli RV (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations. Proc Natl Acad Sci USA 88: 9828–9832
28. Müller-Starck G (1992) Genetic control of inheritance of isoenzymes in poplars of the Tacamahaca section and hybrids. Silvae Genetica 41: 87–95
29. Paterson AH, DeVerna JW, Lanini B and Tanksley SD (1990) Fine mapping of quantitative trait loci using selected overlapping recombinant chromosomes, in an interspecies cross of tomato. Genetics 124: 735–742
30. Pinon J (1992) Variability in the genus *Populus* and sensibility to *Melampsora* rusts. Silvae Genetica 41: 25–33
31. Radetzky R (1990) Analysis of mitochondrial DNA and its inheritance in *Populus*. Curr Genet 18: 429–434
32. Rajora OP (1988) Allozymes as aids for identification and differentiation of some *Populus maximowiczii* Henry clonal varieties. Biochem Syst Ecol 16: 635–640
33. Rajora OP (1989) Characterization of 43 *Populus nigra* L. clones representing selections, cultivars and botanical varieties based on their allozyme genotypes. Euphytica 43: 197–206
34. Rajora OP and Dancik BP (1995) Chloroplast DNA variation in *Populus*. III. Novel chloroplast DNA variants in natural *Populus* × *canadensis* hybrids. Theor Appl Genet 90: 331–334
35. Rajora OP and Dancik BP (1995) Chloroplast DNA variation in *Populus*. I. Intraspecific restriction fragment diversity within *Populus deltoides*, *P. nigra* and *P. maximowiczii*. Theor Appl Genet 90: 317–323
36. Rajora OP and Dancik BP (1995) Chloroplast DNA variation in *Populus*. II. Interspecific restriction fragment polymorphisms and genetic relationships among *Populus deltoides*, *P. nigra*, *P. maximowiczii*, and *P.* × *canadensis*. Theor Appl Genet 90: 324–330
37. Rajora OP and Zsuffa L (1989) Multilocus genetic structure, characterization and relationships of Populus × canadensis cultivars. Genome 32: 99–108
38. Tanksley SD, Bernatzky R, Lapitan NL and Prince JP (1988) Conservation of gene repertoire but gene order in pepper and tomato. Proc Natl Acad Sci USA 85: 6419–6423
39. Tanksley SD, Ganal MW and Martin GB (1995) Chromosome landing: a paradigm for map-based gene cloning in plants with large genomes. Trends Genet 11: 63–68
40. Tanksley SD, Ganal MW, Prince JP, de Vicente MC, Bonierbale MW, Broun P, Fulton TM, Giovannoni JJ, Grandillo S, Martin GB, Messeguer R, Miller JC, Miller L, Paterson AH, Pineda O, Röder MS, Wing RA, Wu W and Young ND (1992)

High density molecular linkage maps of the tomato and potato genomes. Genetics 132: 1141–1160

41. Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, and Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Res 23: 4407–4414
42. Wang J and Hall R (1995) Comparison of DNA content and chromosome numbers in species and clones of poplar. International Poplar Symposium. Seattle, WA, p 93
43. Weeden NF, Muehlbauer FJ and Ladizinsky G (1992) Extensive conservation of linkage relationships between pea and lentil genetic maps. J Hered 83: 123–129
44. Whitkus R, Doebley J and Lee M (1992) Comparative genome mapping of sorghum and maize. Genetics 132: 1119–1130
45. Zabeau M and Vos P (1993) Selective restriction fragment amplification: A general method for DNA fingerprinting. European Patent Application 92402629.7.199333. Publ No 0 534 858 A1

Heredity 84 (2000) 143–151 Received 24 March 1999, accepted 16 August 1999

013

Microsatellite analysis of the regeneration process of *Magnolia obovata* Thunb.

YUJI ISAGI*†, TATSUO KANAZASHI‡, WAJIROU SUZUKI§, HIROSHI TANAKA§ & TETSUTO ABE§

†Kansai Research Centre, Forestry and Forest Products Research Institute, Momoyama, Fushimi, Kyoto 612-0855, Japan, ‡Tohoku Research Centre, Forestry and Forest Products Research Institute, Morioka, Iwate 020-0123, Japan and §Forest Environment Division, Forestry and Forest Products Research Institute, Tsukuba, Ibaraki 305-8687, Japan

We analysed the regeneration process of *Magnolia obovata* using polymorphic microsatellite markers. Eighty-three adult trees standing in a watershed covering an area of 69 ha, and saplings collected from a smaller research plot (6 ha) located at the centre of the watershed were genotyped using microsatellite markers. Among 91 saplings analysed, 24 (26%) had both parents, 31 (34%) had one parent and 36 (40%) had no parent within the watershed. The proportion of genes in saplings inherited from the adults within the watershed was 43%, and therefore 57% were from outside the site, indicating active gene exchange across the watershed area. Average distance between parents and saplings (264.6 ± 135.3 (SD) m) was significantly smaller than that of pairs randomly chosen between adults and saplings (436.7 ± 203.0 (SD) m). The distance of pollen movement inferred from the distance between the two parents of each sapling ranged from 3.2 m to 540 m with an average of 131.1 m ± 121.1 m (SD). Because 34% (=31/91) of saplings had only one parent within the watershed, the estimate of average pollen movement must be smaller than the actual one. Long-distance seed dispersal by birds, inbreeding depression and limitation in acceptance of pollen because of the difference of phenology in each individual flower were considered to be the probable causes of large gene exchange across the watershed.

Keywords: gene flow, pollen dispersal, pollination, seed dispersal.

Introduction

Magnolia obovata is a large, common deciduous tree of temperate forests in Japan reaching 30 m in height. Its large flowers do not secrete nectar, and are primarily pollinated by beetles (Thien, 1974) which are thought to be less efficient than bees. The flowers are protogynous and usually close between the female and male period (Kikuzawa & Mizui, 1990). Although the flowering period of each flower is 3–4 days, for an individual tree flowering persists for up to 40 days (Kikuzawa & Mizui, 1990). The standing density of adult trees is relatively low at a few trees per hectare. In temperate forests in Japan, a few dominant tree species often occupy a large proportion of the canopy, e.g. *Fagus crenata*, and the rest of the canopy is composed of tree species occurring at relatively low density such as *Kalopanax pictus*, *Cornus controversa*, *Aesculus turbinata*, *Magnolia obovata*, *Magnolia salicifolia* and *Pterocarya rhoifolia*. Although each species is at a relatively low density, the assemblage of these species is dominant as a whole, and determines the structure and diversity of the forest ecosystem. For such species, it is important to analyse the extent of pollen movement and seed dispersal to elucidate the regeneration process, mechanisms that maintain biological diversity in forest tree communities and also for conservation purposes.

The pattern and degree of gene dispersal can affect the genetic structure of plant populations (e.g. Schaal, 1980; Ellstrand, 1992; Hamrick *et al.*, 1992), and in higher plants, gene dispersal occurs at reproduction through pollen and seeds. Microsatellite loci are ideal for quantifying pollen- or seed-mediated gene transfer in natural plant populations because of their codominant inheritance and high variability. Therefore, they should provide high exclusion probabilities for paternity assignment. We have developed 11 microsatellite marker loci in *Magnolia obovata* (Isagi *et al.*, 1999) to assign parentage and examine gene transfer of this species. In the present study, we will estimate pollen and seed

*Correspondence. E-mail: isagiy@fsm.affrc.go.jp

dispersal distances and the magnitude of gene transfer in a population where adult tree density was 1.2 ha^{-1}.

Materials and methods

Field site

Field research was conducted in Ogawa Forest Preserve, Ibaraki Pref., central Japan (36°56′N, 140°35′E). The elevation of the research area was 610–660 m a.s.l. and annual mean air temperature and annual precipitation were 9°C and 1800 mm, respectively. Dominant woody species in the canopy were *Quercus serrata*, *Fagus japonica* and *Fagus crenata*, etc. We established two research plots, plots A and B, in the preserve (Fig. 1). Plot A covered the whole watershed area of 69 ha. Within this plot all of the adult trees of *M. obovata* were located and the diameters at breast height (d.b.h.) were measured. The other plot, plot B, occupied an area of 6 ha (200 × 300 m), located at the centre of plot A (Fig. 1). In plot B, intensive studies on the plant community structure, community dynamics (Nakashizuka *et al.*, 1992) and population dynamics of various tree species such as *Carpinus* (Shibata & Nakashizuka, 1995), *Acer* (Tanaka, 1995) and *Cornus* (Masaki *et al.*, 1994) have been made.

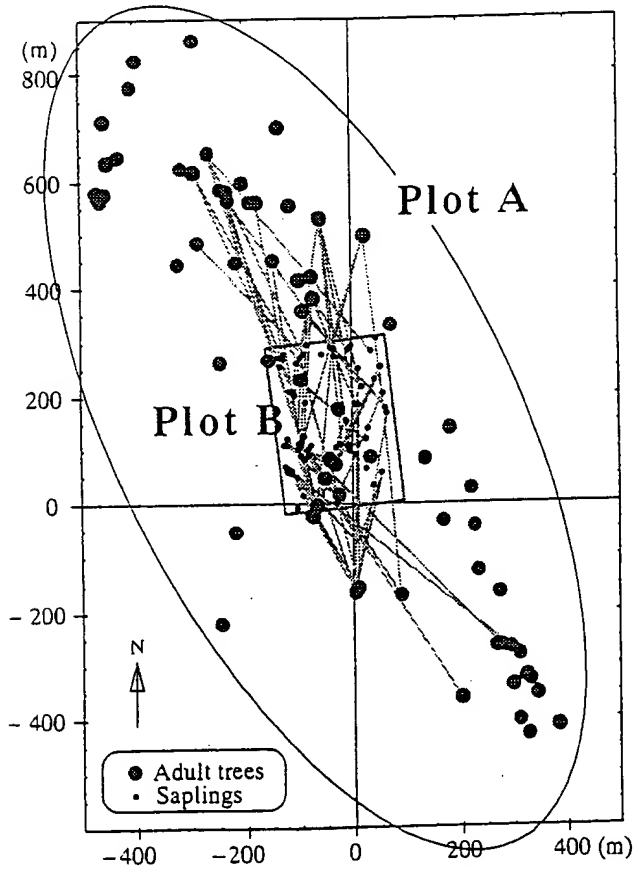


Fig. 1 Map showing the distribution of adults and saplings of *Magnolia obovata* analysed with lines drawn between parents and their offspring.

Sampling

Leaf samples from all of the reproductive adult trees (83 trees) of *M. obovata* in plot A were collected. In plot B, leaf tissue was sampled from 91 saplings. During leaf collection, the position of each tree was mapped. Leaf samples were stored at −70°C prior to DNA extraction.

DNA extraction and microsatellite analysis

Crude genomic DNA of *M. obovata* was extracted using the CTAB method (Milligan, 1992). Genotypes of each DNA sample were scored using eight pairs of microsatellite PCR primers developed by Isagi *et al.* (1999). PCR amplifications were performed, using a thermal cycler (GeneAmp PCR System 9600, ABI), under the following conditions: initial denaturing at 94°C for 9 min, then 30 cycles of denaturation at 94°C for 30 s, annealing for 30 s, and extension at 72°C for 1 min, followed by a final incubation at 72°C for 7 min. The volume of the reaction mixture was 10 μL containing 10 ng of DNA from *M. obovata*, 5 pmol of primers labelled with fluorescent phosphoramidites (TET or 6-FAM), 0.25 U of Taq polymerase (Ampli TaqGold, ABI), 200 μM of each dNTP, 1.5 mM of $MgCl_2$, 10 mM of Tris-HCl, pH 8.3, 50 mM of KCl and 0.001% of gelatin. The PCR products were resolved on a 5% denaturing polyacrylamide gel, and the sizes were determined by automated fluorescent scanning detection with the autosequencer ABI377 and GeneScan™ analysis software (ABI).

Parentage analysis

Parentage was assigned by comparing alleles between a sapling and candidate parents (Dow & Ashley, 1996) using genotype data at eight microsatellite loci: *M6D1*, *M6D3*, *M6D4*, *M10D3*, *M10D6*, *M10D8*, *M15D5* and *M17D5* developed by Isagi *et al.* (1999). Alleles at every locus of each sapling were compared with those of adult trees, and adults which did not share any alleles at each locus were excluded as candidate parents.

Results

Analysis of parentage

The eight microsatellite markers were highly variable, and sufficiently informative to conduct the analysis of

Table 1 Allele frequencies, numbers of heterozygotes and homozygotes, observed and expected heterozygosities at eight microsatellite loci in 83 adults and 91 saplings of *Magnolia obovata*

	Frequency (%)			Frequency (%)			Frequency (%)	
Allele	Adults	Saplings	Allele	Adults	Saplings	Allele	Adults	Saplings
Locus *M6D4*			Locus *M10D3*			Locus *M10D6*		
146		1.1	209		0.6	271	2.4	1.7
148	0.6	0.6	211	6.0	8.2	273	21.1	28.0
150	13.9	14.8	215		1.1	275		1.7
151	1.2		217		1.1	277	12.1	6.6
152	15.1	15.9	219		2.2	279	19.3	23.6
154	4.2	7.1	221	9.0	5.0	281	16.9	9.3
155	0.6	0.6	223	1.8	0.6	283	11.5	10.4
156	19.9	24.2	225	4.8	2.2	285	16.3	18.1
158	4.2	7.7	227	7.2	12.6	287	0.6	0.6
160	2.4	3.3	229	1.8	3.9			
161	0.6		232	1.2	2.2	No. of alleles	8	9
162	0.6	0.6	236	1.2	1.1	Heterozygotes	74	75
164	1.8	1.1	238	21.1	16.5	Homozygotes	9	16
165		0.6	240	8.4	14.8	H_o	0.89	0.82
167	1.2	0.6	242	3.6	1.1	H_e	0.84	0.81
171	0.6		244	3.6	5.5	Locus *M10D8*		
173	1.8	0.6	246	1.8		257		0.6
175		0.6	248	3.0	3.9	267	0.6	1.7
177	0.6	1.1	250	12.7	8.8	275	1.8	
179	1.8	1.7	252	1.8	0.6	277		0.6
181	0.6	0.6	254	1.2		279	16.3	14.8
183	2.4	0.6	258	0.6		281	1.2	2.8
185	1.2	0.6	262	1.8	0.6	282	15.7	25.3
187	0.6	3.3	264	3.0	1.7	283	4.2	1.7
189	3.6	1.7	265	4.2	6.0	284	3.0	2.2
191	1.8					285	0.6	1.1
193	0.6		No. of alleles	21	22	286		0.6
197	3.0	2.2	Heterozygotes	79	88	287	4.2	0.6
199	1.2		Homozygotes	4	3	288	0.6	
201	1.8	2.8	H_o	0.95	0.97	289	10.8	8.2
203		1.1	H_e	0.91	0.91	290	0.6	0.6
205	1.8	1.1	Locus *M15D5*			291	4.2	
210	0.6		96	1.8	1.1	293	6.0	9.9
214	1.8		98	4.2	4.4	295	1.2	1.1
216	3.0	2.8	100	47.0	33.5	297	13.3	8.8
218	1.2	0.6	102	27.1	32.4	299	5.4	7.1
220	1.2		104	17.5	24.7	300		1.1
222	1.2	0.6	106		1.1	301	4.2	3.9
228	0.6		108	2.4	2.8	302	1.8	0.6
235	0.6	0.6				303	3.6	7.1
			No. of alleles	6	7	313	0.6	
No. of alleles	36	30	Heterozygotes	55	64			
Heterozygotes	76	80	Homozygotes	28	27	No. of alleles	21	21
Homozygotes	7	11	H_o	0.66	0.70	Heterozygotes	76	79
H_o	0.92	0.88	H_e	0.68	0.72	Homozygotes	7	12
H_e	0.91	0.88				H_o	0.92	0.87
						H_e	0.91	0.88

Table 1 (*Continued*)

Allele	Frequency (%) Adults	Saplings	Allele	Frequency (%) Adults	Saplings	Allele	Frequency (%) Adults	Saplings
Locus *M6D1*			Locus *M17D5*			Locus *M6D3*		
130	1.2	1.1	289	1.2		106	10.2	7.7
136	0.6		293	18.1	20.9	116	4.8	2.2
138	1.2		295	3.6	3.9	118		1.7
143	3.6	4.4	297	6.6	3.9	120	6.6	9.9
145	0.6		299	27.1	21.4	122	4.2	5.5
147	4.2	1.1	301	1.8	13.2	124	5.4	0.6
149	7.2	10.0	303	12.1	11.0	126	3.0	3.3
151	4.8	5.6	305	12.1	14.8	128	6.6	11.0
153	1.8	4.4	307	7.2	8.2	130	5.4	8.2
155	1.2	1.7	313		2.2	132	4.2	6.6
159	4.8	3.9	317		0.6	134	2.4	1.7
161	1.8					136	3.0	3.3
167	3.6	2.8	No. of alleles	9	10	138	0.0	0.6
169	6.0	5.6	Heterozygotes	65	79	141	2.4	4.4
171	13.9	11.7	Homozygotes	18	12	145	1.2	
173	5.4	3.3	H_o	0.78	0.87	147	0.6	
175	6.6	8.9	H_e	0.84	0.85	150	6.6	4.4
177	4.8	3.9				154	0.6	0.6
179	1.2	1.7				158	0.6	0.6
181	0.6	0.6				160	3.0	5.0
183	2.4	1.7				162	2.4	1.7
185	7.2	4.4				164	0.6	3.3
187	1.8	2.8				166	5.4	2.8
189	2.4	2.2				168	3.6	2.8
191	1.8	0.6				170	7.2	7.1
193	1.8	6.1				172	0.6	0.6
195	5.4	9.4				174	1.8	
197	1.2					178	5.4	4.4
199		1.7				180		0.6
201	0.6					182	0.6	
215	0.6	0.6				192	0.6	
						198	0.6	
No. of alleles	30	25				No. of alleles	29	26
Heterozygotes	80	89				Heterozygotes	81	87
Homozygotes	3	1				Homozygotes	2	4
H_o	0.96	0.99				H_o	0.98	0.96
H_e	0.95	0.94				H_e	0.95	0.94

parentage. The number of alleles for each locus ranged from six (*M15D5*) to 36 (*M6D4*) with an average of 20.3 for adults and from seven (*M15D5*) to 30 (*M6D4*) with an average of 18.8 for saplings (Table 1). The number of alleles unique to adults and saplings was 30 and 19, respectively.

Among 91 saplings found in plot B, 55 had at least one parent (first parent) in plot A whereas 36 had no parents within the watershed. Among the 55 saplings, 24 had the second parent as an exact match, and 31 had only one parent in plot A. Out of the 31 saplings which had only one parent in plot A, 23 had only one candidate as an exact match, and eight had multiple matches with two candidates for one parent. No sapling seemed to be the product of self-pollination of adults in plot A.

In order to estimate the amount of gene flow into the watershed, it is important to evaluate the amount of cryptic gene flow which reflects the possibility that a sapling identified as having a parent within the research plot actually had the parent outside the plot (Dow & Ashley, 1996). Using the allele frequencies at the eight

Table 2 Number of saplings having the first and second parents within or outside plot A, and values of cryptic gene flow for *Magnolia obovata*

	First parent	Second parent	Total
Saplings having parents within plot A	55 (53.95)	24 (23.99)	79 (77.94)
Saplings having parents outside plot A	67 (68.05)	36 (36.01)	103 (104.06)
Cryptic gene flow	1.05	0.01	

Figures in parentheses are numbers of saplings corrected for the values of cryptic gene flow. Because the number of saplings analysed was 91, figures in the column 'Total' for either saplings having parents within or outside plot A should sum up to 182 (=2 parents × 91 saplings).

microsatellite loci and the formula of Marshall *et al.* (1998), the probability of excluding a single randomly chosen unrelated individual in plot A from parentage was determined as 0.999769 for the first parents and 0.999995 for the second parents. Therefore, the probability of excluding correctly all unrelated adults (83 trees) within plot A was $0.999769^{83} = 0.9810$ for the first parents and $0.999995^{83} = 0.9996$ for the second parents. The number of saplings which had the first and second parents in plot A was 55 and 24, respectively (Table 2), so that the amount of cryptic gene flow was estimated as $55(1 - 0.9810) = 1.05$ for the first parent and $24(1 - 0.9996) = 0.01$ for the second parent. Therefore, the total gene flow events from outside plot A into plot B corrected for cryptic gene flow were 68.05 + 36.01 = 104.06 (Table 2). Among 182 possible parents (=91 saplings × 2) of saplings in plot B, 104.06 (57%) were outside plot A, indicating active gene flow across the watershed.

Distance between parents and saplings

Distance between parents and saplings, which represents either seed dispersal from maternal parents, and pollen movement plus seed dispersal from paternal parents, was large (Fig. 1), ranging from 32.4 m to 563.2 m with an average of 264.6 m ± 135.3 m (SD) (Fig. 2a). Although the value indicates active pollen and seed dispersal in the research site, the distance is limited to saplings for which parentage has been assigned within plot A. Therefore, it probably represents an underestimate of the true distance because 57% of the parents of saplings were outside plot A. The distance between random pairs of adults in plot A and saplings in plot B ranged from 10.3 m to 933.8 m with an average of 436.7 m ± 203.0 (SD) m (Fig. 2b), and was significantly greater than that of the distance between offspring and parent trees (*U*-test, $P < 0.0001$) (Fig. 2). This indicates that trees at a closer distance contribute more as parents of the saplings.

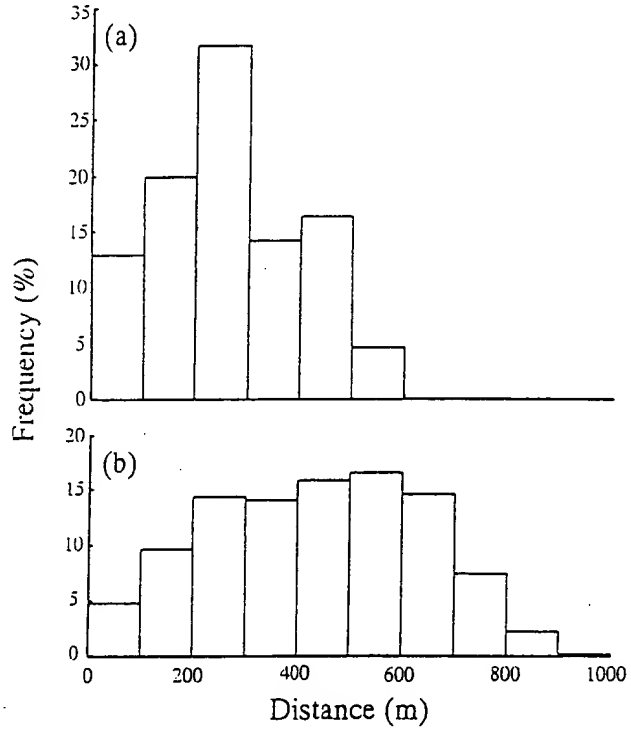


Fig. 2 Histogram of distances for *Magnolia obovata* (a) between parents and progeny inferred with microsatellite markers, and (b) between adult trees and saplings randomly chosen in the research site.

Adult d.b.h.

Diameter at breast height (d.b.h.) of adult trees ranged from 5.2 cm to 59.1 cm with an average of 28.3 cm. The average d.b.h. of adult trees that had progeny in plot B was 35.8 cm, and was significantly larger (*U*-test,

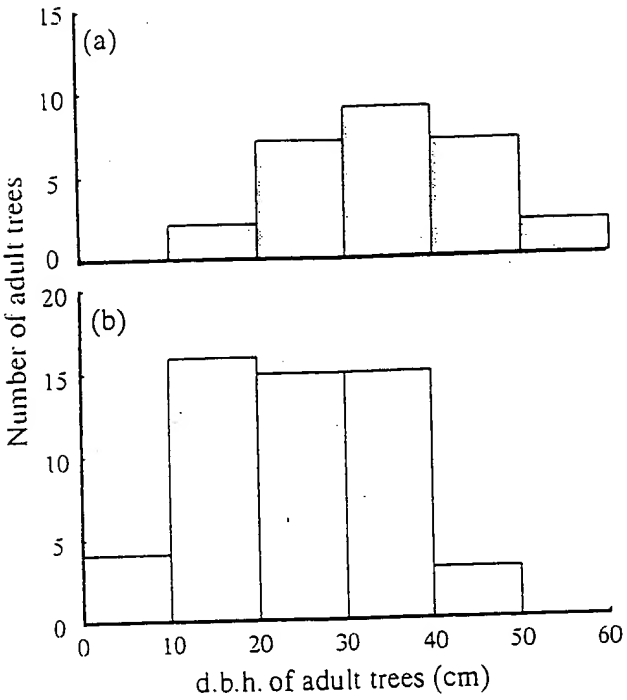


Fig. 3 Diameter at breast height (d.b.h.) of adult *Magnolia obovata*. (a) d.b.h. of adults having their progeny in plot B. (b) d.b.h. of adults not having their progeny in plot B.

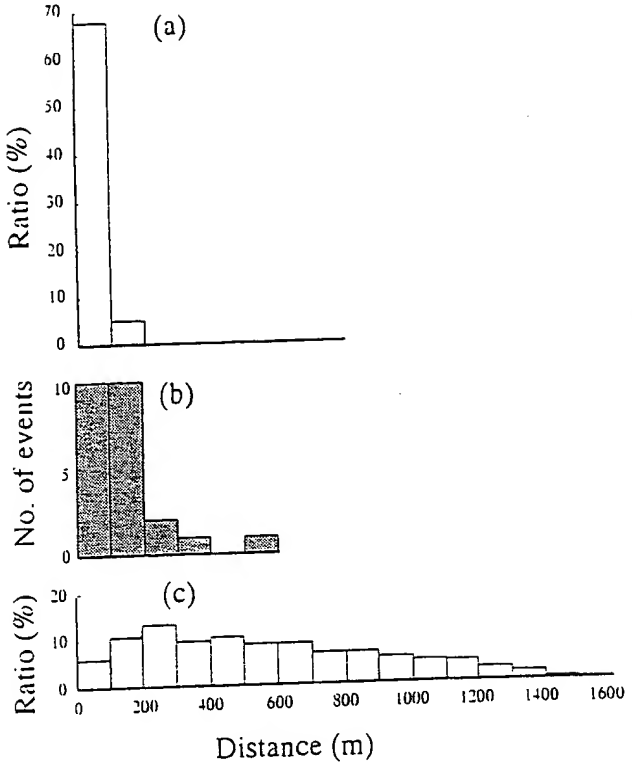


Fig. 4 Histogram for *Magnolia obovata* of distances (a) to nearest neighbours for each adult tree, (b) of pollen movement inferred from microsatellite analysis, and (c) between random pairs of adult trees in plot A.

$P < 0.0001$) than that of adults not having their progeny in plot B (24.5 cm), indicating that adults of larger size contributed more as pollen donors or seed parents to saplings in plot B (Fig. 3).

Pollen movement

Although we determined two parents of exact match for 24 saplings, it was impossible to distinguish which acted as pollen donor or seed parent of these saplings. Therefore, we can not infer the real distance of seed dispersal by merely determining parent–offspring relationships. However, based on the distance between two parents of exact match, we can infer the extent of pollen movement. The distance of pollen movement ranged from 3.2 m to 540 m with an average of 131.1 m ± 121.1 m (SD). About 27% of pollination was carried out between nearest neighbours.

The distances between random pairs of adults in the watershed showed a flat distribution, ranging from 1.3 m to 1543.7 m with an average of 561.5 m (Fig. 4c). The distance between the nearest neighbours for each adult tree within the watershed was low; 93% of trees had their nearest neighbour within the range of 100 m (Fig. 4a). The average distance of pollen movement was significantly larger (*U*-test, $P = 0.0010$) than the average distance between nearest neighbours for each adult tree (44.1 ± 37.5 (SD) m), and significantly smaller (*U*-test, $P < 0.0001$) than that between random pairs of adult trees (561.5 ± 352.6 (SD) m) (Fig. 4c). This indicates that pollination occurs between adults located at closer than the average distance between adult trees within the watershed, but is not always between nearest neighbours (Fig. 5).

The average distance between parents and their offspring (264.6 m; Fig. 2a) was significantly greater (*U*-test, $P < 0.0001$) than that of pollen movement (131.1 m; Fig. 4b), reflecting that the former distance consists of pollen movement and seed dispersal.

Discussion

Pollen dispersal

Movement of pollen grains and seeds from point sources is known to show a leptokurtic or limited distribution (Sork, 1984; Ellstrand, 1992; Webb, 1998): many are dispersed near the source and there is a long tail of fewer

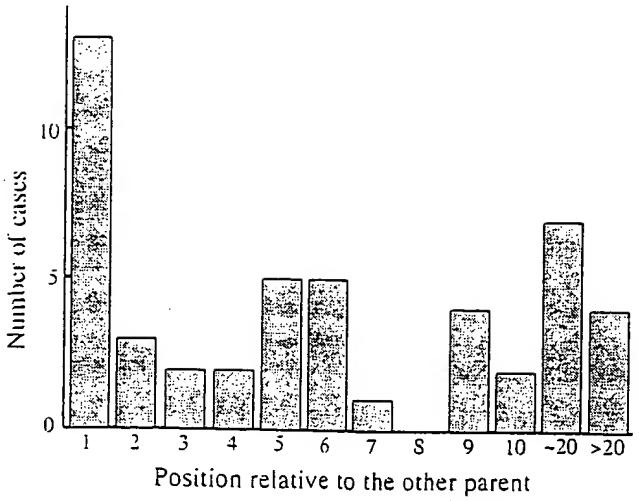


Fig. 5 Relative positions for *Magnolia obovata* of pollen donors and seed parents between which pollination occurred. Position 1 indicates that the tree is the nearest to the other parent among adult trees, and position n indicates that there are $n-1$ trees between the two trees.

pollen grains and seeds at greater distances. However, the distribution of distances between parents and offspring inferred for *M. obovata* with microsatellite markers in the present study was not leptokurtic.

Different pollen vectors and patterns of behaviour remarkably affect pollen dispersal (Schmitt, 1980; Waser, 1982; Hamrick *et al.*, 1992; Webb, 1998). The pollinators for *M. obovata* are primarily beetles (Kikuzawa & Mizui, 1990), which are thought to be less efficient as pollen vectors (Ramsey, 1988). However, the distance of pollen movement in the present stand was quite large (average 131.1 m with a maximum value of 540 m) (Fig. 4b). Tanaka & Yahara (1988) have shown that a variety of insects, other than beetles, i.e. butterflies and bumble bees, pollinated *M. obovata* at a site in central Japan. It is feasible therefore that such pollinators were also effective and account for the long-distance pollen movement in the present stand. About one-third of saplings (=31/91) found in plot B had only one of two parents within plot A. Hence, the average value of pollen movement must be an underestimate and the maximum distance may be more than 540 m. Chase *et al.* (1996) analysed the range of pollen dispersal for a tropical tree species, *Pithecellobium elegans*, using microsatellite markers. They found that average pollen dispersal was 142 m with a maximum value of 350 m. The average distance of pollen movement presently inferred for *M. obovata* is almost equivalent to that for *P. elegans*. Adult trees of *M. obovata* resemble *P. elegans* in that both species tend to occur at low density. It is possible that for tree species occurring at low density in natural communities pollination regularly occurs over a wide range. However, long-distance pollen flow for such tree species might be affected by various life cycle characteristics, habitat type, and type of pollen vectors. This will need to be examined in the future.

Seed dissemination

Seed dispersal characteristics affect the range of seed dissemination (Hamrick & Murawski, 1990), with short distances occurring for gravity dispersal and longer ones for animal and wind dispersal (Hamrick *et al.*, 1992).

Diaspores of *M. obovata* have a red fleshy edible part, and are dispersed internally by birds. Distance of seed dispersal by birds has been considered to be quite long, for example blue jays carried acorns of *Fagus grandifolia* up to 4 km from the source (Johnson & Adkisson, 1985); however, few studies have measured this trait. Instead of direct measurement, the range has been estimated, for example, by determining the home range of birds (e.g. Fukui, 1995). Using appropriate microsatellite markers, and assuming that saplings with no possible parents within the research plot might grow from seeds pollinated outside the research plot and carried in by birds, we can infer the approximate range of seed dispersal by birds. It is notable that 40% (=36/91) of the saplings in plot B had no parents within the 69 ha research site (plot A), and the large proportion of these saplings reflects the active seed dispersal of this species by birds. The range of seed dispersal seems to reach more than several hundred metres and is significantly greater than that reported for seed dispersal by gravity or mammals such as mice and monkeys, whose dispersal ranges are within 100 m from the source (Sork, 1984; Jensen, 1985; Iida, 1996; Yumoto *et al.*, 1998). Dow & Ashley (1996) analysed acorn dispersal of *Quercus macrocarpa* using microsatellite markers, and found that seed dispersal of *Q. macrocarpa* was limited compared with pollen movement; most seed dispersal was less than 30 m whereas pollen dispersal averaged 76.9 m. However, they also stated that long-distance seed dispersal was not so rare as previously estimated: 48% of the seeds were dispersed secondarily by animals, and among them 16% were dispersed more than 90 m away from the source. They also estimated the maximum frequency of saplings which had no parents within the research plot (about 5 ha) at 14%. In contrast, despite the much larger plot size for the present population of *M. obovata* (69 ha), the proportion of saplings without either parent within the research plot was larger (36/91=40%).

Factors that cause large gene transfer for M. obovata

In natural plant populations, it has often been observed that actual gene flow occurs over greater distances than expected allowing for the leptokurtic movement of pollen and seeds. Several factors are thought to account for this discrepancy, namely, the cumulative contribution of pollen by means of leptokurtic but long-tailed distribution (Adams, 1992; Ellstrand, 1992), underestimation of gene dispersal by neglecting carry-over of pollen grains on vectors (Schaal, 1980; Levin, 1981), and inbreeding depression (Waser, 1993).

If a population is genetically structured, and inbreeding or outbreeding depression occurs based on the genetic relatedness of adult trees, some kind of selection on pollen grains could occur. Dow & Ashley (1996) presumed the existence of a mechanism allowing female flowers of *Q. macrocarpa* to select preferentially pollen from distant sources rather than pollen produced by neighbouring trees. It is known that *M. obovata* suffers from high inbreeding depression (Ishida & Nakamura, 1997), and consequently, pollination between less related or spatially distant trees of *M. obovata* might be favoured in spite of the leptokurtic nature of pollen dispersal.

Chase *et al.* (1996) revealed that most mating events in *Pithecellobium elegans* were not between the closest neighbours, because of variation in phenology or in flowering behaviour between adult trees. Many tree species show large fluctuations in flowering among years with or without synchronization between trees in a population (Kelly, 1994; Isagi *et al.*, 1997). In the case of episodic flowering without synchronization in a population, only some of the trees in the population can contribute to reproduction in a given year, and this may result in pollination between distant trees. Mating events in *M. obovata* were also not usually between nearest neighbours: more than 70% of pollination occurred between non-nearest neighbours (Fig. 5), and this might also stem from differences in phenology of individual flowers. Although *M. obovata* has a long flowering period, up to 40 days, the longevity of each protogynous flower is a few days: duration of the consecutive female and male stages is about 1–2 days each. And in most cases, only several or fewer flowers on an adult tree bloom in a given day during the flowering period. Therefore, even within a flowering season, each individual tree may switch among male, female and bisexual phases, and thus not all trees in a population can contribute as pollen donors to flowers in the female stage at the same time.

For *M. obovata* (i) long-distance seed dissemination by birds (ii) inbreeding depression and (iii) limitation in acceptance of pollen for each tree caused by differences in flowering period of each individual flower, are likely to cause long-distance gene flow and increase gene exchange between less related or distant trees. Although the present population of *M. obovata* is in a physically distinct landscape component — a watershed — the amount of gene flow from outside the watershed was sufficient to prevent genetic differentiation by means of genetic drift. This agrees with the fact that most tree species do not exhibit much genetic differentiation among populations: usually more than 90% of the total genetic variation is found within each population (Ledig, 1986).

Acknowledgements

We thank S. Isagi, K. Ishida, H. Ito, Y. Kiyono, T. Nakashizuka, K. Odani and M. Shibata for their help with this work.

References

ADAMS, W. T. 1992. Gene dispersal within forest tree populations. *New Forests*, 6, 217–240.

CHASE, M. R., MOLLER, C., KESSELI, R. AND BAWA, K. S. 1996. Distant gene flow in tropical trees. *Nature*, 383, 398–399.

DOW, B. D. AND ASHLEY, M. V. 1996. Microsatellite analysis of seed dispersal and parentage of saplings in bur oak, *Quercus macrocarpa*. *Mol. Ecol.*, 5, 615–627.

ELLSTRAND, N. C. 1992. Gene flow among seed plant populations. *New Forests*, 6, 241–256.

FUKUI, A. 1995. The role of the brown-eared bulbul *Hypsypetes amaurotis* as a seed dispersal agent. *Res. Pop. Ecol.*, 37, 211–218.

HAMRICK, J. L. AND MURAWSKI, D. A. 1990. The breeding structure of tropical tree populations. *Pl. Sp. Biol.*, 5, 157–165.

HAMRICK, J. L., GODT, M. J. AND SHERMAN-BROYLES, S. L. 1992. Factors influencing levels of genetic diversity in woody plant species. *New Forests*, 6, 95–124.

IIDA, S. 1996. Quantitative analysis of acorn transportation by rodents using magnetic locator. *Vegetatio*, 124, 39–43.

ISAGI, Y., SUGIMURA, K., SUMIDA, A. AND ITO, H. 1997. How does masting happen and synchronize? *J. Theor. Biol.*, 187, 231–239.

ISAGI, Y., KANAZASHI, T., SUZUKI, W., TANAKA, H. AND ABE, T. 1999. Polymorphic DNA markers for *Magnolia obovata* Thunb. and their utility in the related species. *Mol. Ecol.*, 8, 698–700.

ISHIDA, K. AND NAKAMURA, K. 1997. Outcrossing rate and inbreeding depression in natural populations of *Magnolia hypoleuca*. *Am. J. Bot. Abstract*, 84, 121.

JENSEN, T. S. 1985. Seed–seed predator interactions of European beech, *Fagus silvatica* and forest rodents, *Clethrionomys glareolus* and *Apodemus flavicollis*. *Oikos*, 44, 149–156.

JOHNSON, W. C. AND ADKISSON, C. A. 1985. Dispersal of beech nuts by blue jays in fragmented landscapes. *Am. Midl. Nat.*, 113, 319–324.

KELLY, D. 1994. The evolutionary ecology of mast seeding. *Trends Ecol. Evol.*, 9, 465–471.

KIKUZAWA, K. AND MIZUI, N. 1990. Flowering and fruiting phenology of *Magnolia hypoleuca*. *Pl. Sp. Biol.*, 5, 255–261.

LEDIG, F. T. 1986. Heterozygosity, heterosis, and fitness in outbreeding plants. In: Soulé, M. E. (ed.) *Conservation Biology*, pp. 77–104. Sinauer Associates, Sunderland, MA.

LEVIN, D. A. 1981. Dispersal versus gene flow in plants. *Ann. Mo. Bot. Gard.*, 68, 233–253.

MARSHALL, T. C., SLATE, J., KRUUK, L. E. B. AND PEMBERTON, J. M. 1998. Statistical confidence for likelihood-based paternity inference in natural populations. *Mol. Ecol.*, 7, 639–655.

MASAKI, T., KOMINAMI, Y. AND NAKASHIZUKA, T. 1994. Spatial and seasonal patterns of seed dissemination of *Cornus controversa* in a temperate forest. *Ecology*, 75, 1903–1910.

MILLIGAN, B. 1992. Plant DNA isolation. In: Hoezel, A. R. (ed.) *Molecular Genetic Analysis of Populations: A Practical Approach*, pp. 59–88. IRL Press, Oxford.

NAKASHIZUKA, T., IIDA, S., TANAKA, H., SHIBATA, M., ABE, S., MASAKI, T. *ET AL.* 1992. Community dynamics of Ogawa Forest Reserve, a species-rich deciduous forest, central Japan. *Vegetatio*, 103, 105–112.

RAMSEY, M. W. 1988. Differences in pollinator effectiveness of birds and insects visiting *Banksia menziesii* (Proteaceae). *Oecologia*, 76, 119–124.

SCHAAL, B. A. 1980. Measurement of gene flow in *Lupinus texensis*. *Nature*, 284, 450–451.

SCHMITT, J. 1980. Pollinator foraging behavior and gene dispersal in *Senecio* (Compositae). *Evolution*, 34, 934–943.

SHIBATA, M. AND NAKASHIZUKA, T. 1995. Seed and seedling demography of four co-occurring *Carpinus* species in a temperate deciduous forest. *Ecology*, 76, 1099–1108.

SORK, V. L. 1984. Examination of seed dispersal and survival in red oak, *Quercus rubra* (Fagaceae), using metal-tagged acorns. *Ecology*, 65, 1020–1022.

TANAKA, H. 1995. Seed demography of three co-occurring *Acer* species in a Japanese temperate deciduous forest. *J. Veg. Sci.*, 6, 887–896.

TANAKA, H. AND YAHARA, T. 1988. The pollination of *Magnolia obovata*. In: Kawano, S. (ed.) *The World of Plants II*, p. 37. Kenkyusya, Tokyo.

THIEN, L. B. 1974. Floral biology of *Magnolia*. *Am. J. Bot.*, 61, 1037–1045.

WASER, N. M. 1982. A comparison of distances flown by different visitors to flowers of the same species. *Oecologia*, 55, 251–257.

WASER, N. M. 1993. Population structure, optimal outbreeding, and assortative mating in angiosperms. In: Thornhill, N. W. (ed.) *The Natural History of Inbreeding and Outbreeding*, pp. 173–199. The University of Chicago Press, Chicago.

WEBB, C. L. 1998. The selection of pollen and seed dispersal in plants. *Pl. Sp. Biol.*, 13, 57–67.

YUMOTO, T., NOMA, N. AND MARUHASHI, T. 1998. Cheek-pouch dispersal of seeds by Japanese monkeys (*Macaca fuscata yakui*) on Yakushima Island, Japan. *Primates*, 39, 325–338.

Aust. J. Bot., 1998, 46, 533–546

An Evaluation of the AFLP Fingerprinting Technique for the Analysis of Paternity in Natural Populations of *Persoonia mollis* (Proteaceae)

Siegfried L. Krauss[AB] *and Rod Peakall*[A]

[A]Division of Botany and Zoology, The Australian National University, Canberra, ACT 0200, Australia.
[B]Corresponding author; email: Siegy.Krauss@anu.edu.au

Abstract

The accurate assignment of paternity in natural plant populations is required to address important issues in evolutionary biology, such as the factors that affect reproductive success. Newly developed molecular fingerprinting techniques offer the potential to address these aims. Here, we evaluate the utility of a new PCR-based multi-locus fingerprinting technique called Amplified Fragment Length Polymorphism (AFLP) for paternity studies in *Persoonia mollis* (Proteaceae). AFLPs were initially scored for five individuals from three taxonomic levels for 64 primer pairs: between species (*P. mollis* and *P. levis*), between subspecies (*P. mollis* subsp. *nectens* and subsp. *livens*), between individuals within a single population of *P. mollis*, as well as for a naturally pollinated seed from a single *P. mollis* subsp. *nectens* plant. Overall, 1164 fragments (24.6% of all fragments) were polymorphic between species, 743 (16.5%) between subspecies, 371 (8.6%) between individuals within a single population, and 265 (6.2%) between a plant and its seed. Within a single *P. mollis* population of 14 plants, 42 polymorphic fragments were scored from profiles generated by a single AFLP primer pair. The mean frequency of the recessive allele (q) over these 42 loci was 0.773. Based on these observations, it will be feasible to generate well over 100 polymorphic AFLP loci with as few as three AFLP primer pairs. This level of polymorphism is sufficient to assign paternity unambiguously to more than 99% of all seed in experiments involving small, known paternity pools. More generally, the AFLP procedure is well suited to molecular ecological studies, because it produces more polymorphism than allozymes or RAPDs but, unlike conventionally developed microsatellite loci, it requires no prior sequence knowledge and minimal development time.

Introduction

The transmission of genes from one generation to the next by sexual reproduction is a process of major evolutionary significance, as it influences the genetic makeup of future generations. Despite its significance, many fundamental questions remain unanswered. For example, what are the factors that influence reproductive success? In plants, do some pollen donors sire more seeds than others in natural populations? What is the intensity of post-pollination selection? What are the evolutionary consequences of variable reproductive success? Answers to these sorts of questions require the accurate assignment of male parentage to progeny using genetic markers in conjunction with appropriate experimental manipulations of natural populations.

We are conducting experiments to address some of these questions in natural populations of *Persoonia mollis* R.Br. (Proteaceae), a long-lived, fire-sensitive shrub, occurring up to c. 250 km south and up to c. 150 km west of Sydney, New South Wales (NSW), Australia. Currently, nine subspecies are recognised (Krauss and Johnson 1991). Flowers are borne singly within the axils of leaves, and are pollinated by a suite of bees (Bernhardt and Weston 1996). The species is completely outcrossing due to a 'pseudo' self-incompatibility mechanism that prevents almost all self-pollen tube growth in the style, coupled with preferential development of outcrossed seed (Krauss 1994*a*, *b*). Ovaries contain two ovules, only one of which usually matures into seed. Since individual stigmas of *P. mollis* usually receive mixed pollen loads (Krauss 1994*a*), and, given that there is usually only one (male

0067-1924/98/030533

genotype) winner per ovary, there is enormous potential for post-pollination selection through male–male competition among pollen and/or female choice, both of which can affect male reproductive success (Snow 1994). To measure the intensity of post-pollination selection, we are conducting pollination-manipulation experiments in a small natural population of *P. mollis* subsp. *nectens* at Sublime Point, NSW. Final results of this experiment will be presented in detail elsewhere, once seeds have been assigned paternity. Our purpose here is to describe the genetic methodology and to evaluate its utility for paternity analysis in this system.

Allozymes have been used to assign paternity, although only rarely in natural plant populations (e.g. Ellstrand and Marshall 1986; Adams *et al.* 1992). In *P. mollis*, for example, only 61 of 940 (6.5%) seeds genotyped in natural populations could be assigned paternity unambiguously using allozymes (Krauss 1994*a*). Similarly, in *Chamaelirium luteum*, only 55 of 2255 (2.4%) seeds could be assigned paternity unambiguously (Smouse and Meagher 1994). Except in special circumstances, for example where plants have singly-sired 'clutches' (e.g. milkweeds, Broyles and Wyatt 1990), allozymes are (at best) only sufficiently variable to partially assign paternity (Chakraborty *et al.* 1988). Clearly, more variable genetic markers are required. Below, we briefly introduce some of the new molecular techniques that may enable unambiguous paternity assignment in natural plant populations, before evaluating one of these methods (AFLPs) for this purpose.

Brief Background to Molecular Markers

The revolutionary advent of the polymerase chain reaction (PCR) (Mullis and Faloona 1987; Saiki *et al.* 1988) has enabled the development of new methods that overcome the limitations of allozymes and RFLP methods, and in theory allowed access to a vast array of genetic markers. The PCR is a procedure that can exponentially replicate either single-locus or multi-locus DNA, resulting in up to a million-fold amplification (Arnheim *et al.* 1990). The PCR reaction consists of template DNA, DNA polymerase, two synthesised primers (typically 8–30 nucleotides in length) that anneal to complementary regions of template DNA, and other additives in a buffer. This reaction mix is subjected to repeated temperature cycles in a thermocycler. First the mix is heated to 94°C to separate the double-stranded DNA. Next the temperature is lowered (to between 40°C and 60°C) to allow the primers to anneal to their complementary sequences. Finally, the temperature is raised to 72°C to allow the addition of nucleotides between the flanking primers by the polymerase. The new double-stranded DNA fragments serve as the starting point for the next cycle, and fragment copy number thus increases exponentially. From small amounts of starting material, PCR can yield sufficient DNA for visualisation on an electrophoretic gel. Specific fragments that differ in length can be readily detected as genetic markers that vary both within and among individuals (Arnheim *et al.* 1990).

The many variations of the basic PCR procedure can be broadly defined by three major approaches. Firstly, sequence-tagged-site (STS) PCR uses two different specific primers, complementary to opposite strands of conserved DNA flanking regions, to amplify the intervening sequences. Microsatellites, or simple-sequence-repeats (SSRs), are a type of STS marker. SSRs consist of tandemly repeated units of a short nucleotide motif one to six base pairs long. Dinucleotide repeats (e.g. CACACA ...), trinucloetide repeats (e.g. AATAATAAT ...), and tetranucleotide repeats (e.g. GATAGATAGATA ...) are the most common, and are widely distributed throughout the genomes of eukaryotes (Jarne and Lagoda 1996). Different alleles show different numbers of repeat units, which manifest as length polymorphisms on a gel. Being co-dominant markers, both heterozygotes and homozygotes can be detected at single loci. The utility of SSRs results from their inherent multi-allelic variability. In humans, for example, heterozygosities typically exceed 50%, with

up to 50 alleles per locus (Weber 1990). Similar patterns of variability at SSR loci have been found in mammals, birds, insects and plants (Queller *et al.* 1993; Chase *et al.* 1996; Gupta *et al.* 1996; Koreth *et al.* 1996; Primmer *et al.* 1996). By virtue of their variability and co-dominance, SSRs are considered ideal genetic markers for a wide range of applications including paternity analysis. The major disadvantage of SSRs is that sequences on either side of the SSRs must be known to design appropriate primers for PCR assay. Thus, a substantial investment in time and cost is required to develop these markers via the construction of genomic libraries and DNA sequencing.

A second PCR approach overcomes the need for sequence knowledge, but brings its own limitations. Williams *et al.* (1990) and Welsh and McClelland (1990) showed that by using a single short primer of known but arbitrary sequence under low-stringency PCR conditions, polymorphic multi-locus DNA profiles can be produced. Although other names exist for this procedure, it is most widely known as Random Amplified Polymorphic DNA (RAPDs). In general, RAPDs are dominant with polymorphism revealed as band presence or absence. Typically, 5–20 bands are produced per PCR reaction, with each band assumed to represent a single locus. While dominance reduces the information content per locus, RAPDs tend to detect more variation than allozymes, because many more loci can be assayed (Peakall *et al.* 1995). Concerns have been raised, however, about run-to-run repeatability and PCR artefacts which may occur as a consequence of the low-stringency conditions used (Morell *et al.* 1995).

A third PCR approach combines the features of both the STS and RAPD techniques. The Amplified Fragment Length Polymorphism (AFLP) procedure is based on selective PCR amplification of particular restriction fragments from a total digest of genomic DNA (Lin and Kuo 1995; Vos *et al.* 1995). The basic method involves restriction of the genomic DNA, ligation of oligonucleotide adapters to the DNA fragments, and high-stringency selective amplification of a subset of all the fragments in the total digest. The ligation of oligonucleotide adapters enables PCR to be performed for any species without prior sequence knowledge. The selective amplification uses primers of complementary sequence to the ligated adapter plus one to three additional arbitrary nucleotides. Subsequent electrophoresis of the PCR product typically reveals a complex multi-locus profile of up to 100 bands. The bands are generally dominant markers, with polymorphism detected as band presence or absence. The AFLP method has a number of significant advantages over RAPDs. First and foremost, the PCR conditions are far more stringent and optimised for plants, producing highly reproducible profiles. When combined with fluorescently labelled primers, electrophoresis on an automated sequencer and computer data storage and analysis, the power, speed and accuracy of this procedure is unsurpassed for a multi-locus fingerprinting technique.

Although a new technique, AFLPs have already been used to assess genetic relationships, to quantify levels of genetic diversity, and to identify cultivars of agriculturally important plants (e.g. Hill *et al.* 1996; Maughan *et al.* 1996; Powell *et al.* 1996; Sharma *et al.* 1996; Donini *et al.* 1997; Qi and Lindhout 1997; Van Toai *et al.* 1997). In this paper, we consider the utility of AFLPs for ecological genetic studies in natural plant populations. Specifically, we evaluate polymorphism generated by AFLP for paternity assignment in natural populations of *P. mollis*.

Materials and Methods

DNA Extraction

Genomic DNA was isolated from c. 30 mg of fresh new leaf using a modified CTAB procedure described by Stewart and Via (1993), and as suggested for AFLP by Vos *et al.* (1995), but with the addition of a phenol–chloroform step. DNA was similarly extracted from a naturally pollinated seed harvested from one plant in the study population. DNA samples were stored in low Tris-EDTA (TE) buffer at –20°C.

Amplified Fragment Length Polymorphism (AFLP)

The AFLP procedure involves four steps (Fig. 1).

(*i*) Restriction of the DNA. For each sample, approximately 200 ng of DNA was digested with 2.5 units of EcoRI/MseI restriction enzyme in a reaction volume of 25 µL, and incubated at 37°C for 2 h. Samples were then transferred to a 70°C bath for 15 min, before briefly cooling on ice.

(*ii*) Ligation of adapters. One unit of DNA ligase and 24 µL of adapter ligation solution were added to the digested DNA from (i), incubated at 20°C for 2 h, then diluted 1:10 with TE buffer.

(*iii*) Preselective amplification by PCR. Five µL of the diluted ligation mix was combined with 40 µl pre-amplification primer solution, 5 µl 10X PCR buffer for AFLP, and 2.5 units *Taq* DNA polymerase in a PCR plate and PCR was performed for 20 cycles of 94°C for 30 s, 56°C for 60 s, and 72°C for 60 s. Subsequently, the pre-amplification mixture was diluted 1 : 50 with TE buffer.

(*iv*) Selective amplification by PCR. This step uses primers that match the known adapter sequence, plus three selective nucleotides, to reduce the complexity of the profile. For each of 64 primer pairs (Table 1), the following was added for 2.5 µL of each diluted, pre-selective DNA sample from (iii): 7.5 ng EcoRI-primer, 15 ng MseI-primer, 3.95 µL MilliQ water, 1.0 µL 10X PCR buffer, and 0.25 units *Taq* DNA polymerase. A touchdown PCR reaction commenced with one cycle of 94°C for 30 s, 65°C for 30 s, and 72°C for 60 s. In subsequent cycles, the annealing temperature was reduced in 1°C steps to 56°C, followed by 23 cycles at 56°C. All solutions, except for Taq DNA polymerase, come in kit form and were purchased from Life Technologies. Fluorescently labelled primers were purchased from Perkin Elmer. PCR was performed on a Corbett Research FTS-960 thermal sequencer.

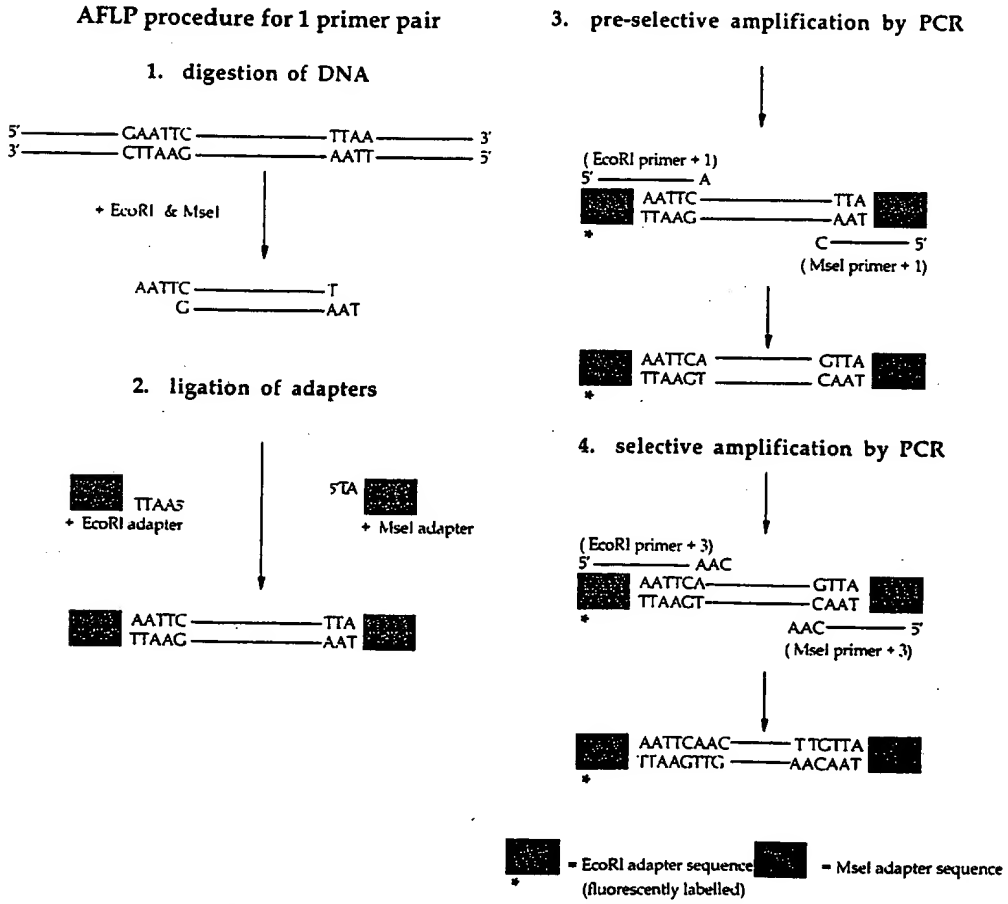


Fig. 1. The four major steps of the AFLP procedure. See text for a detailed explanation.

Table 1. AFLP selective amplification primers used here to generate 64 AFLP primer pairs

Fluorescently labelled primers	Unlabelled primers
EcoRI-ACT	MseI-CAA
EcoRI-ACA	MseI-CAC
EcoRI-AAC	MseI-CAG
EcoRI-ACC	MseI-CAT
EcoRI-AGC	MseI-CTA
EcoRI-AAG	MseI-CTC
EcoRI-AGG	MseI-CTG
EcoRI-ACG	MseI-CTT

The fluorescently labelled amplified fragments were analysed by gel electrophoresis (5% acrylamide gels), using the ABI Prism 377 Automated Genetic Analysis System (AGAS). In this system, fragments are detected by laser, and are accurately sized by the inclusion of internal size standards that are labelled with a uniquely coloured fluorescent dye. Laser technology and fluorescent labels permit more efficient electrophoresis, allowing visualisation of up to three different PCR products (with different coloured fluorescent primers) per lane, with the fourth colour devoted to the size standard. Digitally converted raw data are saved on the computer as samples migrate past the fluorescence detector. Multi-locus profiles were visualised using ABI GeneScan software. AFLP profiles were scored for presence or absence of fragments, as well as for fragment intensity, as revealed by peak height, with the aid of ABI Genotyper software.

To maximise the efficiency of this preliminary evaluation, we first contrasted AFLP profiles for 64 primer pairs for each of five individuals for four hierarchical levels of comparison. The contrasts were between a single plant of *P. mollis* subsp. *nectens* from Sublime Point and (i) its seed, (ii) a co-occurring plant, (iii) *P. mollis* subsp. *livens*, and (iv) *P. levis*. To test the power of the AFLP method to distinguish between genotypes within populations, we chose two individuals of *P. mollis* at Sublime Point that had identical allozyme phenotypes (Krauss 1997). Replicate extractions and AFLP analyses were made for each plant. The number of fragments of lengths between 80 and 500 bases, the number of polymorphic fragments, and the percentage of all fragments that were polymorphic, were scored for each primer pair. Secondly, a more detailed investigation was made for one of the most polymorphic AFLP primer pairs for 14 plants within the Sublime Point population of *P. mollis* subsp. *nectens*. The number of polymorphic fragments and the frequency of each of these fragments were scored.

Results

A portion of one multi-locus profile, as created by GeneScan software, for each of two co-occurring individuals of *P. mollis* subsp. *nectens*, shows five clear polymorphic fragments distinguishing these individuals (Fig. 2). The quality of individual multi-locus profiles was occasionally affected by poor PCR reactions or by variable gel conditions, such as bubbles in the gel matrix and sub-standard wells for loading samples. However, these factors reduced peak heights (and therefore yielded a weaker profile), rather than changing the quantitative nature of the profile (presence or absence of peaks) in replicate samples. Poor profiles were re-run.

On average, approximately 70 fragments were scored per primer pair. In total, 64 AFLP primer pairs generated 4722 fragments, of which 1164 (24.6%) were polymorphic for the inter-species comparison between one plant of *P. mollis* and one plant of *P. levis* (Table 2). For the inter-subspecies comparison between one plant of *P. mollis* subsp. *nectens* and one plant of subsp. *livens*, 743 fragments (16.5%) were polymorphic from a total of 4509 fragments. From a total of 4323 fragments, 371 (8.6%) were polymorphic for the within-population comparison between two plants of *P. mollis* subsp. *nectens* at Sublime Point. The mother–naturally pollinated seed comparison generated 4279 fragments, 265 (6.2%) of which were polymorphic (Table 2).

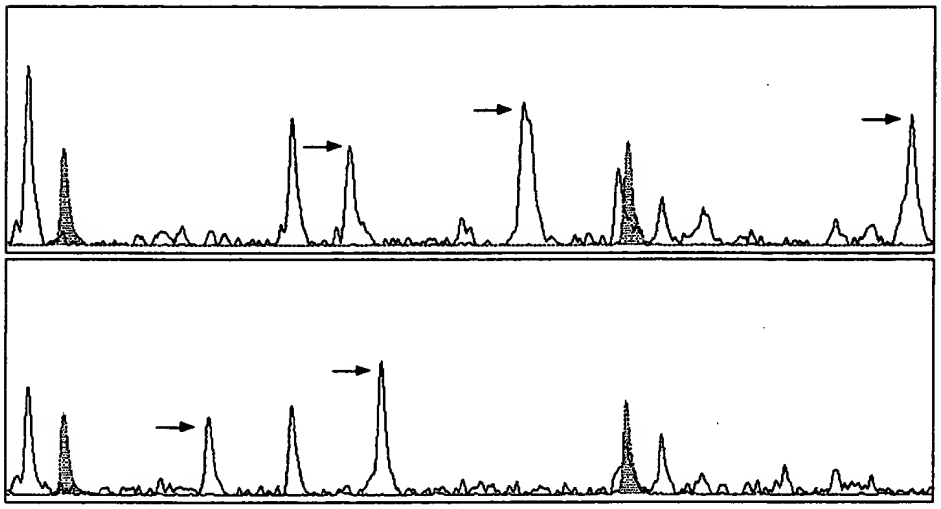


Fig. 2. A portion of the AFLP profile as displayed by GeneScan for two co-occurring plants of *P. mollis* subsp. *nectens*. Shown is the profile for M-CAC and E-ACG for fragments between 195 and 280 bases in length. The size standards (200 and 250 bases) are shaded peaks. Note the polymorphisms at 213, 225, 228, 241 and 275 bases (arrowed).

Polymorphism varied markedly for different primer pairs. For example, one of the most polymorphic primer pairs (M-CAG and E-ACA) produced 42 polymorphic fragments (52% of all fragments) between species, 25 (33%) between subspecies, 15 (28%) between individuals within the one population of *P. mollis*, and 10 (15.6%) between a maternal plant and one of its seeds. The least variable primer pair (M-CAT and E-AAG) produced no polymorphism (Table 2). The diminishing marginal information return, when adding primer pairs in order of decreasing levels of polymorphism, is slight for the species and subspecies comparison, and more pronounced for the within-population and mother–offspring comparisons (Fig. 3), highlighting the greater difficulty of finding markers for paternity analysis. Thus, in our data set, 40 primer pairs account for almost all polymorphic fragments between two individuals of *P. mollis* co-occurring within a single population, and the seven most polymorphic primer pairs generated over 100 polymorphic fragments.

The most informative primer pairs from the inter-species comparison were M-CTG and E-ACG (28 polymorphic fragments, 33% of all fragments), M-CAG and E-ACA (42 fragments, 52%), M-CTG and E-ACT (40 fragments, 34%), and M-CAC and E-ACT (34 fragments, 30%) (Table 2). The most informative primer pairs from the inter-subspecies comparison were M-CTG and E-ACT (29 fragments, 27%), M-CAG and E-ACA (25 fragments, 33%), M-CTA and E-AGG (25 fragments, 24%), and M-CAC and E-ACT (23 fragments, 21%). The most informative primer pairs from the within-population comparison were M-CTC and E-ACG (16 fragments, 19 %), M-CAG and E-ACA (15 fragments, 23%), M-CTA and E-AGG (15 fragments, 15%), M-CTG and E-AGG (15 fragments, 17%) and M-CAC/E-ACG (13 fragments, 20%). The most informative primer pairs from the mother/seed comparison were M-CTC and E-ACG (13 fragments, 16%), M-CTG and E-AGG (13 fragments,14%), M-CTG and E-ACT (11 fragments,10%) and M-CAC and E-ACG (10 fragments, 15%) (Table 2). Therefore, different sets of primer pairs may be the most informative at different taxonomic levels.

In a population of 14 plants of *P. mollis* subsp. *nectens* at Sublime Point, NSW, 42 polymorphic fragments were scored from profiles generated by the AFLP primer pair

Table 2. Number of polymorphic fragments (# polys), total number of fragments (# frags) and percentage of all fragments polymorphic (% polys) for each of 64 AFLP primer pair between one individual of *P. mollis* subsp. *nectens* and (i) one individual of *P. levis* (between species), (ii) one individual of *P. mollis* subsp. *livens* (between subspecies), (iii) one co-occurring individual of *P. mollis* subsp. *nectens* (within population), and (iv) one progeny (seed)

		x-species			x-subsp			within pop			mother/seed		
		# frags	# polys	% polys	# frags	# polys	% polys	# frags	# polys	% polys	# frags	# polys	% polys
m-primer	e-primer												
CAA	AAC (Y)	56	14	25	56	10	17.8	52	4	7.7	50	2	4
	AAG (G)	64	11	17.2	71	10	14.1	62	3	4.8	62	2	3.2
	ACA (B)	63	21	33.3	56	10	17.9	51	2	3.9	50	2	4
	ACC (Y)	54	20	37	54	13	24.1	52	5	9.6	48	2	4.2
	ACG (G)	46	15	32.6	46	12	26.1	43	4	9.3	40	1	2.5
	ACT (B)	99	23	23.2	99	14	14.1	85	8	9.4	84	3	3.6
	AGC (Y)	45	9	20	42	3	7.1	39	1	2.6	39	2	5.1
	AGG (G)	104	18	17.3	89	8	9	92	7	7.6	90	4	4.4
CAC	AAC (Y)	38	3	7.9	36	3	8.3	36	0	0	36	0	0
	AAG (G)	91	23	25.3	85	12	14.1	84	5	5.9	83	2	2.4
	ACA (B)	81	11	13.6	75	7	9.3	73	4	5.5	73	6	8.2
	ACC (Y)	91	25	27.5	94	21	22.3	84	12	14.3	87	9	10.3
	ACG (G)	77	18	23.3	66	13	19.7	66	13	19.7	67	10	14.9
	ACT (B)	112	34	30.4	111	23	20.7	104	10	10.4	101	7	6.9
	AGC (Y)	49	10	20.4	48	7	14.6	49	4	8.2	48	1	2.1
	AGG (G)	97	32	33	96	20	20.8	88	7	7.9	85	5	5.9
CAG	AAC (Y)	74	24	32.4	74	14	18.9	65	7	10.7	65	3	4.6
	AAG (G)	82	11	13.4	84	8	9.5	82	6	7.3	78	2	2.5
	ACA (B)	81	42	51.9	75	25	33.3	66	15	22.7	64	10	15.6
	ACC (Y)	119	26	21.8	106	11	9.6	110	7	6.4	107	4	3.7
	ACG (G)	63	32	50.8	60	20	33.3	55	8	14.5	51	1	2
	ACT (B)	127	32	25.2	109	22	20.2	102	7	6.9	100	4	4
	AGC (Y)	91	21	23.1	82	11	13.4	86	7	8.1	79	4	5.1
	AGG (G)	97	17	17.5	92	9	9.8	93	6	6.5	93	9	9.7
CAT	AAC (Y)	38	2	5.3	38	2	5.3	37	1	2.7	37	1	2.7
	AAG (G)	40	0	0	40	0	0	40	0	0	40	0	0
	ACA (B)	90	27	30	89	14	15.7	83	5	6	83	2	2.4
	ACC (Y)	42	18	42.8	36	8	22.2	37	6	16.2	34	5	14.7
	ACG (G)	63	13	20.6	64	14	21.9	59	5	8.5	59	3	5.1
	ACT (B)	73	10	13.7	70	4	5.7	71	2	2.8	74	2	2.7
	AGC (Y)	78	23	29.5	75	17	22.7	68	6	8.8	68	6	8.8
	AGG (G)	64	13	20.3	57	3	5.3	56	1	1.8	54	1	1.8
CTA	AAC (Y)	44	2	4.5	44	4	9.1	43	1	2.3	42	1	2.4
	AAG (G)	83	20	24.1	79	15	19	76	9	11.8	75	10	13.3
	ACA (B)	84	11	13.1	77	7	9.1	77	2	2.6	75	2	2.7
	ACC (Y)	71	12	16.9	67	12	17.9	60	3	5	62	3	4.8
	ACG (G)	69	23	33.3	66	18	27.3	58	7	12.1	61	4	6.6
	ACT (B)	93	18	19.4	92	13	14.1	85	4	4.7	86	3	3.5
	AGC (Y)	68	7	10.3	67	8	11.9	69	7	10.1	69	7	10.1
	AGG (G)	109	28	25.7	106	25	23.6	102	15	14.7	99	8	8.1

Table 2. (*continued*)

x-species		x-subsp			within pop			mother/seed					
		# frags	# polys	% polys	# frags	# polys	% polys	# frags	# polys	% polys	# frags	# polys	% polys
CTC	AAC (Y)	92	22	23.9	85	9	10.6	83	6	7.2	81	5	6.2
	AAG (G)	68	5	7.3	69	5	7.2	65	0	0	69	3	4.3
	ACA (B)	114	19	16.7	111	12	10.8	108	9	8.3	112	6	5.3
	ACC (Y)	56	4	7.1	56	4	7.1	55	3	5.5	54	2	3.7
	ACG (G)	78	28	35.9	81	22	27.2	84	16	19	80	13	16.2
	ACT (B)	71	23	32.4	62	13	20.9	54	6	11.1	55	5	9.1
	AGC (Y)	55	16	29.1	54	12	22.2	51	5	9.8	47	2	4.2
	AGG (G)	100	23	23	88	7	7.9	90	5	5.6	90	3	3.3
CTG	AAC (Y)	69	8	11.6	72	7	9.7	61	2	3.1	61	1	1.6
	AAG (G)	60	18	30	57	9	15.8	57	3	5.3	57	4	7
	ACA (B)	73	22	30.1	70	19	27.1	69	8	11.6	71	8	11.3
	ACC (Y)	59	6	10.1	59	3	5.1	59	3	5.1	58	3	5.2
	ACG (G)	69	28	40.6	59	8	13.6	59	5	8.5	59	4	6.8
	ACT (B)	118	40	33.9	109	29	26.6	108	15	13.9	110	11	10
	AGC (Y)	48	9	18.7	47	11	23.4	43	0	0	43	1	2.3
	AGG (G)	95	27	28.4	91	18	19.8	90	15	16.7	90	13	14.4
CTT	AAC (Y)	25	4	16	22	1	4.5	22	1	4.5	22	1	4.5
	AAG (G)	44	7	15.9	41	4	9.7	40	0	0	41	1	2.4
	ACA (B)	71	24	33.8	70	16	22.9	72	10	13.9	67	6	9
	ACC (Y)	90	29	32.2	86	16	18.6	78	10	12.8	77	4	5.2
	ACG (G)	61	23	37.7	50	6	12	53	6	11.3	53	6	11.3
	ACT (B)	65	20	30.7	65	16	24.6	60	6	10	61	5	8.2
	AGC (Y)	38	18	47.4	38	9	23.7	33	3	9.1	32	0	0
	AGG (G)	93	22	23.7	94	17	18.1	89	8	9	91	5	5.5
	Total 4722	1164	24.65	4509	743	16.48	4323	371	8.58	4279	265	6.19	

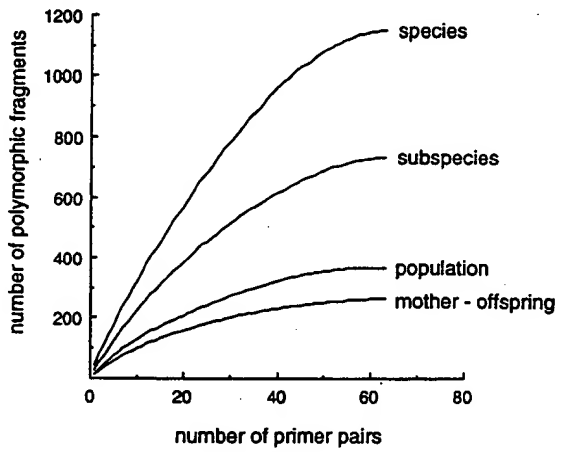


Fig. 3. Plot of cumulative number of polymorphic fragments against number of AFLP primer pairs (from most to least polymorphic), for each of four comparisons: a single plant of *P. mollis* subsp. *nectens* from Sublime Point and (i) its seed ('mother–offspring'), (ii) a co-occurring plant ('population'), (iii) a single plant of *P. mollis* subsp. *livens* ('subspecies'), and (iv) a single plant of *P. levis* ('species').

M-CAC and E-ACG. The frequency of each of these fragments within the population varied from 0.07 (i.e. the fragment unique in the population) to 0.93 (i.e. the fragment possessed by all but one plant) (Fig. 4). The mean frequency of the recessive phenotype (absence of a fragment, q^2) over all 42 fragments was 0.598, and the mean frequency of the recessive allele (q), assuming Hardy–Weinberg equilibrium, was 0.773 (s.e. = 0.07).

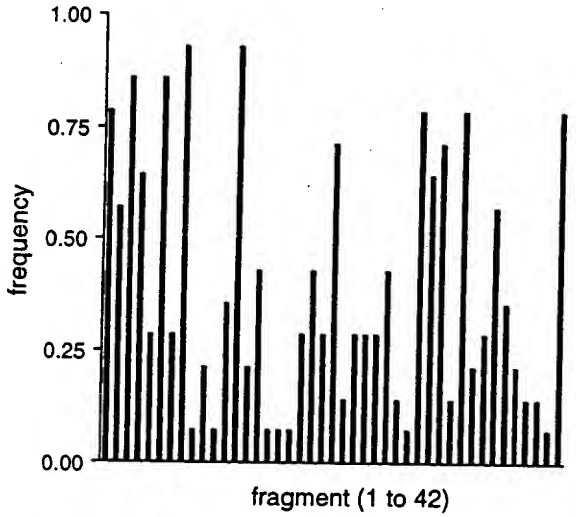


Fig. 4. The frequency of each of 42 polymorphic fragments generated by the AFLP primer pair M-CAC and E-ACG for 14 plants in a natural population of *Persoonia mollis* subsp. *nectens* at Sublime Point, NSW.

Discussion

AFLP for Paternity Analysis

Are AFLPs sufficiently variable to assign paternity unambiguously in natural populations of *P. mollis*? Using the formula described in Lewis and Snow (1992, p. 156), we estimated the fraction of offspring for which paternity can be assigned unambiguously to a single individual (P_{ET}). The variables included the number of plants contributing to the pollen pool, the frequency of the recessive allele (assumed to be constant over all loci), and the number of polymorphic loci assayed. In the pollination experiment being conducted by us to address the intensity of post-pollination selection, 12 out of 14 possible fathers needed to be excluded (the self-male was excluded, as *P. mollis* is essentially self-incompatible). Preliminary studies of population variation generated by one AFLP primer pair (M-CAC and E-ACG) yielded an estimated recessive phenotype frequency per locus (q^2) of 0.598, and an estimated recessive allele frequency (q) of 0.773. Given these estimated frequencies, 50 loci are expected to exclude all but one potential father for 88.0% of all seeds, 75 loci will allow unambiguous paternity assignment of 98.1% of all seeds, while 100 loci will account for 99.7% of all seeds in the study population. This preliminary study has already produced over 42 polymorphic loci from just one AFLP primer pair for a population of 14 plants. It is therefore likely that well over 100 polymorphic loci will ultimately be generated using as few as three AFLP primer pairs.

These estimates of P_{ET} involve a number of assumptions (Lewis and Snow 1992), including, as calculated here, that all loci have a recessive allele frequency of 0.77. As more

polymorphic loci are identified, we will preferentially be able to use those loci with a recessive allele frequency greater than 0.77, increasing the power of P_{ET}. A second assumption is that the potential fathers in the pollen pool are not closely related. The paternity exclusion probability (P_{ET}) is exaggerated when close relatives (e.g. sibs) compete for paternity (Double *et al.* 1997). In plants, the level of consanguineous matings in natural populations can be estimated using the effective selfing model of Ritland and Jain (1981). This model separates the estimate of selfing into two components; true selfing and effective selfing, due to matings with close relatives. Effective selfing will inflate the estimate of selfing. *Persoonia mollis* is consistently completely outcrossing (Krauss 1994*a*), so selfing (and therefore effective selfing) is negligible. Thus, population genetic structure of *P. mollis*, at the scale of the paternity pool (Levin 1988), does not lead to consanguineous matings, probably due to high seed dispersal, high pollen dispersal and extremely high seed mortality.

Advantages of AFLPs

Where allozyme analysis produced identical phenotypes for two co-occurring individuals of *P. mollis* at Sublime Point (Krauss 1997), the 64 AFLP primer pairs produced 371 polymorphic fragments to distinguish between them. At the population level, where 14 alleles were detected at 11 co-dominant allozyme loci (Krauss 1997), 42 dominant loci (each with two alleles, presence or absence) were scored for one AFLP primer pair. Thus, the AFLP method has the power to distinguish between genotypes, even in populations where allozymes show very low levels of genetic diversity. Consequently, AFLP may be particularly useful for ecological genetic studies within the Proteaceae, which generally appears to show unusually low levels of allozyme variability (Krauss 1997). Similarly, AFLP produces many more polymorphic loci per primer than RFLPs, SSRs and RAPDs in the predominantly selfing *Glycine max* (Maughan *et al.* 1996).

Being a PCR technique, the AFLP method can routinely provide a high throughput. After an initial (time-consuming) screening period, where the best (i.e. most polymorphic) primer pairs are identified by laboriously scoring profiles for all primer pairs, the scoring of individual polymorphic fragments for presence or absence can be achieved rapidly. Using the protocol set out in this study, we envisage running and scoring 100 individuals for 100 polymorphic loci per week. Thus, the AFLPS method is ideally suited to ecological genetic studies, where many samples are required.

The use of fluorescently labelled primers, and fluorescent detection using an automatic sequencer and GeneScan software, has a number of major advantages over radioactive labelling. First, the safety concerns associated with radioactive chemicals are avoided. Second, the output is available at the end of a gel run rather than waiting many hours for radioactive incubation. Third, digitally modified data are saved onto the computer as fragments pass the detector and scoring of electropherograms is more efficient and accurate than scoring autoradiographs by eye, particularly as a size-standard is included in each lane for accurate sizing of all fragments. Ultimately, scoring of polymorphic fragments will be automated, using software developed specifically for this system. Fourth, by using differently coloured fluorescent primers, one can multiplex PCR products in a single lane, reducing the number of gels required and making the system cost effective per polymorphic fragment. For example, by combining PCR products from the three most polymorphic primer pairs, it should ultimately be possible to generate over 100 polymorphic loci for small natural populations of *P. mollis* in only one gel lane.

Limitations of AFLPs

One problem requiring further work is that of rare disappearing fragments, especially where fragments are apparently polymorphic. For example, for the primer pair M-CAA and E-AAG, we scored an apparently clear polymorphic fragment of 158 bases in length. However, a

subsequent replicate run (fresh extraction, repeat AFLP protocol) failed to produce this fragment. This could be due to incomplete initial digestion of DNA, perhaps the most common source of artifactual polymorphism (Lin and Kuo 1995), poor amplification of this fragment during PCR, or contamination of DNA. Rare disappearing fragments have been highlighted elsewhere (Stanfield *et al.* 1996), and spermidine has been suggested as an addition to the digestion mixture (Bloch and Grossmann 1987) to alleviate the problem. We have yet to test this in our laboratory. On one occasion, we generated a surprisingly variable profile for one seed that was clearly the product of incomplete digestion, either due to poor DNA quality or insufficient restriction enzymes. Extreme care is needed at this stage of the procedure, and it is essential to confirm that all digestion has been complete before ligation. However, AFLPs appear to be robust to variation in DNA concentration, and Lin and Kuo (1995) found no differences in profiles over a concentration range of 100 ng to 5 µg of genomic DNA.

In our study, the AFLP method produces dominant bi-allelic markers, scored as band presence or absence. Although dominant markers are not nearly as efficient as co-dominant markers for this sort of analysis (Lewis and Snow 1992), this problem is largely offset for AFLP by the very large number of fragments generated (potentially over 100 polymorphic fragments per lane). Note, though, that AFLP markers can potentially be co-dominant, particularly where fragments contain SSRs. However, finding the complementary alleles will on most occasions be very difficult without appropriate crosses and analysis of F_1 and F_2 generations.

One downside to AFLP analysis is the cost per sample. The most expensive step is the ligation of adapters onto restricted fragments, which is common to both radioactive and fluorescent methods. This step costs about $A5 per sample using the available kits and recommended amounts of mix components. Fluorescent primers, while expensive to purchase, are cost effective per analysis and much cheaper than radioactive labels. Ultimately, the cost per sample will depend on the number of primer pairs required for the assignment of paternity to each seed. We estimate that if three primer pairs are required to produce sufficient numbers of polymorphic loci for assignment of paternity, and can be multiplexed at three per lane, the cost per sample will be about $A10 for commercially available kits from Life Technologies or Perkin Elmer. The cost can be further reduced by combining the AFLP selective amplification for three primer pairs (differently labelled) into the one, rather than three, PCR reaction well. This reduces the amount of reagents needed, and has been successfully performed with other species in our laboratory.

A major limitation of fluorescent AFLPs for most researchers will be ready access to an Automated Genetic Analysis System. However, the number of laboratories with dedicated access is increasing, as are the number of commercial laboratories that will perform the electrophoresis.

Beyond AFLP

The AFLP method is of value for studies where highly polymorphic genetic markers are required because it provides a stepping stone to the access of highly variable SSR loci, with potentially reduced development costs. A new technique, called SAMPL (Selective Amplification of Microsatellite Polymorphic Loci), uses the AFLP procedure as a starting point to find SSR loci within AFLP-generated fragments (Witsenboer *et al.* 1997). In the final PCR amplification, a fluorescently labelled SSR-anchored primer (e.g. 5′–CCCGTGTGTGTGTGT–3′) is combined with one of the two AFLP primers to amplify AFLP-generated fragments that contain SSRs. Alternatively, inter-SSR regions can be targeted within the AFLP generated fragments, by using a primer with a 5′ SSR sequence (e.g. 5′–AGAGAGAGAGAGAGAGT–3′), in combination with one of the two AFLP primers for the final PCR amplification. In theory, this method has the potential to visualise co-dominant SSR loci within a multilocus profile.

We have performed preliminary studies of the SAMPL procedure on *P. mollis* with promising results. In a comparison of three individuals of *P. mollis* from the Sublime Point

population, using nine SAMPL primers, we have generated multiple fragments and putative polymorphisms. If required, these fragments can be sequenced to confirm that they contain SSRs. Alternatively, progeny arrays can be run to enable the delineation of co-dominant loci.

Conclusion

This study has demonstrated that the AFLP method produces large numbers of polymorphic fragments within small natural populations of *P. mollis* and when contrasting two individuals from (i) two *Persoonia* species, (ii) two *P. mollis* subspecies, (iii) the same *P. mollis* subsp. *nectens* population, and (iv) a *P. mollis* plant and its seed. More specifically, the AFLP method produces sufficient polymorphism for the potentially unambiguous assignment of paternity in natural populations of *P. mollis*. In combination with appropriate experimental manipulations of natural populations, the AFLP method will allow us to address intriguing, and previously intractable, issues in evolutionary biology, such as the intensity of post-pollination selection and its effect on male reproductive success. More generally, because the AFLP method can be applied to any species, it can be recommended as an extremely powerful tool for ecological genetic studies and studies of genetic diversity within and among natural plant populations, particularly where other markers such as allozymes show little or no variation.

Acknowledgments

We thank Wes Keys for his technical expertise in the laboratory, Peter Smouse and Peter O'Hanlon for valuable comments on the manuscript, and Tim Lloyd for access to the field site. S.K. was supported by an ARC postdoctoral fellowship.

References

Adams, W. T., Griffin, A. R., and Moran, G. F. (1992). Using paternity analysis to measure effective pollen dispersal in plant populations. *The American Naturalist* **140**, 762–780.

Arnheim, N., White, T., and Rainey, W. E. (1990). Application of PCR: organismal and population biology. *Bioscience* **40**, 174–182.

Bernhardt, P., and Weston, P. H. (1996). The pollination ecology of *Persoonia* (Proteaceae). *Telopea* **6**, 775–804.

Bloch, K. D., and Grossmann, B. (1987). Enzymatic manipulation of DNA and RNA. I. Restriction Endonucleases. *Current Protocols in Molecular Biology* **1**, 3.1.1–3.1.8.

Broyles, S. B., and Wyatt, R. (1990). Paternity analysis in a natural population of *Asclepias exaltata*: multiple paternity, functional gender, and the 'pollen donation hypothesis'. *Evolution* **44**, 1454–1468.

Chakraborty, R., Meagher, T. R., and Smouse, P. E. (1988). Parentage analysis with genetic markers in natural populations. I. The expected proportion of offspring with unambiguous paternity. *Genetics* **118**, 527–536.

Chase, M., Kesseli, R., and Bawa, K. (1996). Microsatellite markers for population and conservation genetics of tropical trees. *American Journal of Botany* **83**, 51–57.

Donini, P., Elias, M. L., Bougourd, S. M., and Koebner, R. M. D. (1997). AFLP fingerprinting reveals pattern differences between template DNA extracted from different plant organs. *Genome* **40**, 521–526.

Double, M., Cockburn, A., Barry, S. C., and Smouse, P. E. (1997). Exclusion probabilities for single-locus paternity analysis when related males compete for matings. *Molecular Ecology* **6**, 1155–1166

Ellstrand, N. C., and Marshall, D. L. (1986). Patterns of multiple paternity in populations of *Raphanus sativus*. *Evolution* **40**, 837–842.

Gupta, P. K., Balyan, I. S., Sharma, P. C., and Ramesh, B. (1996). Microsatellites in plants—a new class of molecular markers. *Current Science* **70**, 45–54.

Hill, M., Witsenboer, H., Zabeau, M., Vos, P., Kesseli, R., and Michelmore, R. (1996). PCR-based fingerprinting using AFLPs as a tool for studying genetic relationships in Lactuca spp. *Theoretical and Applied Genetics* **93**, 1202–1210.

Jarne, P., and Lagoda, P. J. L. (1996). Microsatellites, from molecules to populations and back. *Trends in Ecology and Evolution* **11**, 424–429.

Koreth, J., Oleary, J. J., and McGee, J. O. (1996). Microsatellites and PCR genomic analysis. *Journal of Pathology* **178**, 239–248.

Krauss, S. L. (1994*a*). Restricted gene flow within the morphologically complex species *Persoonia mollis* (Proteaceae): contrasting evidence from the mating system and pollen dispersal. *Heredity* **73**, 142–154.

Krauss, S. L. (1994*b*). Preferential outcrossing in the complex species *Persoonia mollis* (Proteaceae). *Oecologia* **97**, 256–264.

Krauss, S. L. (1997). Low genetic diversity in *Persoonia mollis* (Proteaceae), a fire-sensitive shrub occurring in a fire-prone habitat. *Heredity* **78**, 41–49.

Krauss, S. L., and Johnson, L. A. S. (1991). A revision of the complex species *Persoonia mollis* (Proteaceae). *Telopea* **4**, 185–199.

Levin, D. A. (1988). The paternity pool of plants. *The American Naturalist* **132**, 309–317.

Lewis, P. O., and Snow, A. A. (1992). Deterministic paternity exclusion using RAPD markers. *Molecular Ecology* **1**, 155–160.

Lin, J. J., and Kuo, J. (1995). AFLP™: a novel PCR-based assay for plant and bacterial DNA fingerprinting. *Focus* **17**, 66–70.

Maughan, P. J., Maroof, M. A. S., Buss, G. R., and Huestis, G. M. (1996). Amplified Fragment Length Polymorphism (AFLP) in soybean species diversity, inheritance, and near-isogenic line analysis. *Theoretical and Applied Genetics* **93**, 392–401.

Morell, M. K., Peakall, R., Appels, R., Preston, L. R., and Lloyd, H. L. (1995). DNA profiling techniques for plant variety identification. *Australian Journal of Experimental Agriculture* **35**, 807–819.

Mullis, K., and Faloona, F. (1987). Specific synthesis of DNA in vitro via a polymerase catalysed chain reaction. *Methods in Enzymology* **55**, 335–350.

Peakall, R., Smouse, P. E., and Huff, D. R. (1995). Evolutionary implications of allozyme and RAPD variation in diploid populations of Buffalograss. (*Buchloe dactyloides* (Nutt. Engelm.). *Molecular Ecology* **4**, 135–147.

Powell, W., Morgante, M., Andre, C., Hanafey, M., Vogel, J., Tingey, S., and Rafalski, A. (1996). The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Molecular Breeding* **2**, 225–238.

Primmer, C. R., Moller, A. P., and Ellegren, H. (1996). A wide-range survey of cross-species microsatellite amplification in birds. *Molecular Ecology* **5**, 365–378.

Qi, X., and Lindhout, P. (1997). Development of AFLP markers in barley. *Molecular and General Genetics* **254**, 330–336.

Queller, D. C., Strassmann, J. E., and Hughes, C. R. (1993). Microsatellites and kinship. *Trends in Ecology and Evolution* **8**, 285–288.

Ritland, K., and Jain, S. K. (1981). A model for the estimation of outcrossing rate and gene frequencies using *n* independent loci. *Heredity* **47**, 35–52.

Saiki, R. K., Gelfund, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., and Erlich, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487–491.

Sharma, S. K., Knox, M. R., and Ellis, T. H. N. (1996). AFLP analysis of the diversity and phylogeny of *Lens* and its comparison with RAPD analysis. *Theoretical and Applied Genetics* **93**, 751–758.

Smouse, P. E., and Meagher, T. R. (1994). Genetic analysis of male reproductive contributions in *Chamaelirium luteum* (L.) Gray (Liliaceae). *Genetics* **136**, 313–322.

Snow, A. A. (1994). Postpollination selection and male fitness in plants. *The American Naturalist* **144** (**suppl**), S69–S83.

Stanfield, A., Rikkerink, E., and Putterill, J. (1996). Locating AFLP markers linked to the resistance gene *Vf* in apple (abstract). In '4th International DNA Fingerprinting Conference (abstracts)'. p. 55.

Stewart, C. N., and Via, L. E. (1993). A rapid CTAB DNA isolation technique useful for RAPD fingerprinting and other PCR applications. *BioTechniques* **14**, 748–750.

Van Toai, T. T., Peng, J., and St Martin, S. K. (1997). Using AFLP markers to determine the genomic contribution of parents to populations. *Crop Science* **37**, 1370–1373.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M., and Zabeau, M. (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* **23**, 4407–4414.

Weber, J. L. (1990). Informativeness of human (dC-dA)n (dG-dT)n polymorphisms. *Genomics* **7**, 524–530.

Welsh, J., and McClelland, M. (1990). Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acids Research* **18**, 7213–7218.

Williams, J. G. K., Kubelik, A. R., Livak, K. J., Rafalski, J. A., and Tingey, S. V. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research* **18**, 6531–6535.

Witsenboer, H., Vogel J., and Michelmore R. W. (1997). Identification, genetic localization, and allelic diversity of selectively amplified microsatellite polymorphic loci in lettuce and wild relatives (*Lactuca* spp.). *Genome* **40**, 923–936.

Manuscript received 14 May 1997, accepted 26 March 1998
http://www.publish.csiro.au/journals/ajb

Euphytica **101**: 199–206, 1998.

Molecular characterization and genetic relatedness among walnut (*Juglans regia* L.) genotypes based on RAPD markers

F.P. Nicese[1], J.I. Hormaza[2] & G.H. McGranahan
Department of Pomology, University of California, Davis CA 95616, U.S.A.; [1] *present address: Dipartimento di Ortoflorofrutticoltura, via G. Donizetti 6, 50144 Firenze, Italy;* [2] *present address: Unidad de Fruticultura, SIA-DGA, Campus de Aula Dei, Apdo. 727, 50080 Zaragoza, Spain*

Received 18 March 1997; accepted 21 October 1997

Key words: cultivar identification, fingerprinting, genetic similarity, *Juglans regia*, RAPD, walnut

Summary

The potential use of the Randomly Amplified Polymorphic DNA (RAPD) technique for characterization and assessment of genetic relationships was investigated in nineteen walnut (*Juglans regia* L.) genotypes used as parents or released as cultivars from the breeding program of the University of California at Davis. Most of the 72 decamer primers used yielded scorable amplification patterns based on discernable bands. The results obtained produced a unique fingerprint for each of the walnut genotypes studied. Cluster analysis separated the 19 walnut genotypes into two main groups whose differences were related to their pedigree. Genotypes sharing common parents tend to group together and with at least one of the parents. Thus, RAPD markers can detect enough polymorphism to differentiate among walnut genotypes, even among closely related genotypes, and the genetic similarity based on RAPDs appears to reflect the known pedigree information. RAPD technology can be useful in current walnut breeding programs, allowing the identification of new cultivars as well as the assessment of the genetic similarity among genotypes which will help in selecting the best parents to obtain new genetic combinations.

Introduction

The family Juglandaceae consists of seven genera comprising about 60 monoecious tree species. The genus *Juglans* contains about 20 species, all producing edible nuts. Among those, the English or Persian walnut (*Juglans regia* L.) is the most widely cultivated species (McGranahan & Leslie, 1990). Persian walnuts have been grown in California since the day of early Spanish missions. In the last 50 years the genetic base of this crop has been enriched with introductions from Asia and Europe, primarily France (Forde & McGranahan, 1996). All the commercial cultivars currently grown in California can be considered descendants, at least partially, from those gene pools and have originated either as chance seedlings, or from the breeding program of the University of California (Serr, 1969). This breeding program, based on crosses between late season and laterally fruiting genotypes, has released 15 walnut cultivars (Tulecke & McGranahan, 1994). The characterization of these genotypes is important for designing objective and repeatable criteria to protect breeders' right in newly developed cultivars.

Accurate and rapid cultivar identification is especially important in vegetatively propagated plant species such as most fruit trees both for practical breeding purposes and for proprietary rights protection. Unfortunately, the traditional methods for characterization and assessment of genetic variability in perennial fruit crop species, based on morphological, physiological and biochemical studies, are both time consuming and affected by the environment. The introduction of molecular biology techniques, such as DNA-based markers, provides an opportunity for genetic characterization that allows direct comparison of different genetic material independent of environmental influences (Weising et al., 1995).

Initial molecular studies in walnut were carried out using isozymes to assess the inheritance of some enzyme systems (Arulsekar et al., 1986; Aleta et

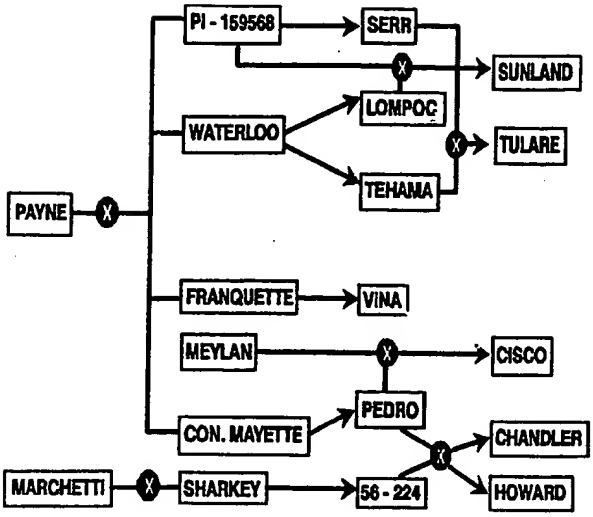


Figure 1. Pedigree diagram of the cultivars tested.

al., 1993). Isozymes have also been used in *Juglans* to identify genetic variability (Malvolti et al., 1993, 1994), to detect interspecific hybrids (Arulsekar et al., 1985; McGranahan et al., 1986; Germain et al., 1993), to identify species and cultivars (Louskas et al., 1984; Wenheng, 1984; Arulsekar et al., 1985; Aleta et al., 1989; Germain et al., 1993; Solar et al., 1993, 1994) and to assess mating parameters (Rink et al., 1994). RFLP markers have also been used in walnuts to determine parentage (Aly et al., 1992), to establish phylogenetic relationships in the genus *Juglans* (Fjellstrom & Parfitt, 1995), to estimate genetic diversity, and to identify cultivars (Fjellstrom et al., 1994; Fjellstrom & Parfitt, 1994a, b). Recently, RAPD markers have been used to evaluate the level of polymorphism at the interspecific level between Persian walnut (*J. regia*) and Northern California black walnut (*J. hindsii* (Jeps.) Jeps.) (Woeste et al., 1996a) and to identify a marker linked to hypersensitivity to the cherry leafroll virus (Woeste et al., 1996b).

Randomly Amplified Polymorphic DNA (RAPD) (Williams et al., 1990, 1993) can be of further use to identify closely related walnut cultivars and could complement the results previously obtained with RFLPs, mainly because of the higher level of polymorphism obtained with RAPDs. RAPD analysis has already proven to be valuable in genotype characterization as well as in population and pedigree analyses in many crop species and studies in tree crops are starting to produce interesting results (Hormaza et al., 1994; Fabbri et al., 1995). RAPD markers are not only important for the characterization of the germplasm but can also be used to evaluate the effects of selection over time and to aid in the development of crossing schemes in walnut improvement programs since this method allows the study of the genetic diversity of the available germplasm.

Table 1. Walnut genotypes included in this study

Code	Genotype	Original source[1]
1	56-224	Univ. of California
2	Chandler	Univ. of California
3	Cisco	Univ. of California
4	Conway Mayette	France
5	Franquette	France
6	Howard	Univ. of California
7	Lompoc	Univ. of California
8	Marchetti	California
9	Meylan	France
10	Payne	California
11	Pedro	Univ. of California
12	PI-159568	Afghanistan
13	Serr	Univ. of California
14	Sharkey	China
15	Sunland	Univ. of California
16	Tehama	Univ. of California
17	Tulare	Univ. of California
18	Vina	Univ. of California
19	Waterloo	California

[1] Based on Tulecke & McGranahan, 1994.

The main objective of this study was to develop RAPD markers to unequivocally characterize 19 closely related walnut genotypes from the breeding program of the University of California as well as to compare the degree of genetic relatedness obtained through RAPD analysis with that of the expected results obtained from pedigree data.

Materials and methods

Plant material

The nineteen genotypes used in this study (Table 1 and Figure 1) were obtained from the walnut collection maintained at the University of California Wolfskill Experimental Orchard in Winters, California, USA, and are part of the walnut breeding program developed at this institution. They were selected based on a pedigree assessment and include recently released cultivars as well as their parental genotypes.

DNA isolation

Young leaves from nineteen accessions of *J. regia* were collected in spring and immediately stored at - 70 °C prior to DNA extraction. Total DNA was extracted following the method of Doyle & Doyle (1987) with minor modifications. Seven grams of leaf tissue were ground to fine powder in liquid nitrogen and added to 20 ml of 65 °C preheated CTAB buffer (2% CTAB, 1% PVP, 1% β-mercaptoethanol, 0.1% sodium bisulfite, 1.4 M NaCl, 100 mM Tris-HCl pH 8.0, 20 mM sodium EDTA) and incubated at 65 °C for 60 min. The lysate was extracted with 20 ml of chloroform/isoamyl alcohol (24 : 1) and centrifuged for 10 min at 1800 rpm in a desktop centrifuge. In order to precipitate the nucleic acids, the aqueous fraction was mixed with an equal volume of cold isopropanol. The nucleic acid precipitate was recovered with a glass hook, washed in 76% ethanol with 10 mM ammonium acetate and air dried overnight before being resuspended in 1 ml TE buffer (10 mM Tris-HCl pH 8.0, 1 mM disodium EDTA). The resuspended DNA was treated with RNAase and the concentration of extracted DNA was determined using a spectrophotometer at 260 nm. DNA was diluted to 10 ng/μl and used for PCR amplification.

DNA amplification and electrophoresis conditions

PCR amplification reactions were carried out as described in Williams et al. (1990) with minor modifications. Reaction mixtures (25 μl total volume) consisted of 50 ng of template DNA, 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.9 mM $MgCl_2$, 0.001% gelatin, 100 μM each of dATP, dGTP, dCTP and dTTP (Perkin-Elmer-Cetus, Norwolk, Conn.), 0.4 M primer (Operon Technologies, Alameda, Calif.) and 0.75 units of Taq DNA polymerase (Perkin-Elmer-Cetus, Norwolk, Conn.; Promega, Madison, Wisc.; Life Technologies, Gaithersburg, Maryl.) overlaid with 25 μl of mineral oil. DNA amplification was carried out in a DNA Thermal Cycler 480 (Perkin-Elmer-Cetus, Norwolk, Conn.) programmed for 1 cycle of 2 min at 94 °C followed by 40 cycles of 45 sec at 94 °C, 1 min at 38 °C and 2 min at 72 °C. After a final incubation for 5 min at 72 °C the samples were stored at 4 °C prior to analysis. The PCR amplified products were resolved by gel electrophoresis in 2.2% Sea Kem agarose (FMC, Rockland, Maine), in TBE buffer (45 mM Tris, 45 mM H_3BO_4, 1 mM EDTA). Gels were stained in ethidium bromide (0.5 μl/ml) and then visualized on a long-wave UV light source.

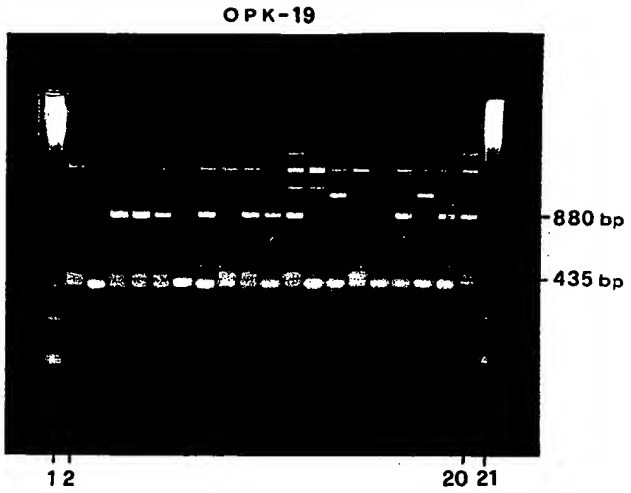


Figure 2. Gel electrophoresis of OPK-19 primer. Lane 1 and 20: 123 bp DNA ladder. Lane 2–19: cultivars tested.

Data analysis

Amplified bands were visually scored as present or absent. Each amplification fragment useful for discrimination between genotypes was named by the source of the primer (OP = Operon), the kit letter, the primer number and its approximate size in base pairs. A similarity matrix was generated using the Nei and Li similarity index (Nei & Li, 1979; Lamboy, 1994) based on the proportion of shared amplification fragments between two genotypes according to the following equation:

$$\text{Similarity} = 2Nab/(Na + Nb)$$

where Nab is the number of scored amplification fragments with the same molecular weight shared between genotypes 'a' and 'b'; Na is the number of scored amplification fragment in genotype 'a'; and Nb is the number of scored amplification fragments in genotype 'b'.

A dendrogram was constructed based on the similarity matrix data by applying unweighted pair group method with arithmetic averages (UPGMA) cluster analysis using the NTSYS-pc computer program version 1.70 (Exeter Software, Setauket, New York).

Results and discussion

A total of 72 decamer primers were used to amplify DNA extracted from the nineteen walnut genotypes used in this study. Almost all the primers yielded scorable amplification patterns (Figure 2). Some of

Table 2. Distribution of 23 RAPD markers within the 19 walnut genotypes. Genotype numbers correspond to those in Table 1. '+' indicates presence and '-' indicates absence of the marker

Markers	Genotypes																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
K01-445	-	+	-	+	-	-	+	-	-	-	+	-	-	-	-	+	+	+	+
K02-280	+	+	+	-	+	-	+	+	+	+	+	+	+	-	+	-	+	+	+
K02-513	-	-	+	-	+	-	+	-	+	-	-	-	-	-	+	+	+	+	+
K03-635	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-
K19-435	+	-	+	+	+	-	-	+	+	-	+	-	-	+	-	-	-	-	+
K19-880	-	-	+	+	+	-	+	-	+	+	+	-	-	-	-	+	-	+	+
K20-1100	+	+	-	+	-	+	-	-	-	-	+	-	-	+	-	-	-	-	-
G10-560	-	+	-	+	+	-	-	-	-	-	+	-	-	+	-	-	-	-	-
G15-690	+	+	-	-	-	-	-	-	-	+	+	+	+	+	-	+	+	+	-
G16-620	-	-	-	+	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-
G16-700	-	-	+	+	-	-	-	-	-	-	+	-	-	-	+	-	-	-	-
G18-630	+	+	-	-	-	+	-	+	-	+	-	+	+	+	-	-	+	+	-
G18-990	+	+	-	-	-	+	-	-	-	-	-	+	-	+	+	-	-	-	-
R06-345	+	+	+	+	-	+	-	+	+	-	+	+	+	-	+	-	+	-	-
R13-430	+	+	+	+	+	-	-	+	+	-	-	+	+	-	+	-	+	-	-
R14-1130	+	+	+	+	+	+	+	+	+	+	+	-	+	-	+	+	+	+	+
R15-490	+	-	-	+	-	-	-	+	+	-	-	-	+	+	-	-	+	-	+
R19-755	-	-	-	-	-	-	-	-	-	-	-	+	+	+	-	-	-	-	-
S01-1550	+	+	-	+	-	+	+	+	+	+	-	+	+	+	+	+	+	-	+
S10-700	+	+	+	+	+	+	+	+	-	+	+	-	-	-	-	+	+	+	+
T04- 630	+	-	-	+	+	-	+	-	-	+	-	+	+	+	+	-	-	-	+
T04-780	-	+	-	+	-	-	-	-	+	-	+	-	-	-	-	-	-	-	-
T06-850	+	+	-	-	-	+	-	-	-	-	-	+	-	+	+	-	-	-	-

the primers, however, produced either no amplification or unreadable gel smears. Eighteen primers, producing 1–2 polymorphic fragments each, generated polymorphic banding patterns among the genotypes studied (Table 2). The total number of polymorphic bands obtained was 23. The apparent low level of polymorphism detected (about 25% of the primers tested) can be explained by the strict criterion adopted to score the markers. Only the conspicuous intensely stained bands between 250 bp and 1700 bp long were considered for analysis. Each RAPD analysis was repeated in separate experiments at least twice, and only hihgly reproducible markers were considered.

Some questions have been raised about the reliability of RAPD data due to their variable nature under different experimental conditions and by the fact that comigrating bands from different individuals do not necessarily represent homologous amplification products (Newbury & Ford-Lloyd, 1993; Bachmann, 1994). However, fragment size can be considered a reliable predictor of homology among closely related individuals, as is the case in this study, although this is not necessarily true at higher taxonomic levels (Rieseberg, 1996). In order to maximize the reliability of the process, the reproducibility of the results obtained was tested in two different ways. First, the amplification patterns obtained with three different Taq DNA polymerases was verified using the same genotype and primer (data not shown). Second, amplification reactions with DNA obtained from different accessions of four genotypes ('Chandler', 'Cisco', 'Howard' and 'Franquette') using two different primers were also compared (data not shown). In both cases, the pattern of amplification was fully reproducible.

The results obtained, using eighteen primers that yield 23 polymorphic RAPD bands (Table 2), produced a unique fingerprint for each of the 19 walnut genotypes included in this study (Table 1) allowing a unequivocal identification of each genotype. Besides, the fingerprint of each genotype is defined by multiple RAPD bands presumably at multiple genetic loci. This is important for cultivar characterization since each cultivar is not defined by a single marker but by a set of several markers. This high level of polymorphism

Table 3. *Table 3.*Similarity matrix generated using the Nei and Li's index. Cultivar numbers correspond to those in Table 1

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	1.00																		
2	0.79	1.00																	
3	0.52	0.43	1.00																
4	0.62	0.62	0.58	1.00															
5	0.52	0.43	0.78	0.58	1.00														
6	0.73	0.73	0.35	0.43	0.23	1.00													
7	0.45	0.45	0.59	0.52	0.71	0.37	1.00												
8	0.78	0.61	0.67	0.58	0.56	0.59	0.47	1.00											
9	0.58	0.50	0.74	0.64	0.63	0.33	0.56	0.74	1.00										
10	0.64	0.54	0.47	0.43	0.59	0.50	0.75	0.59	0.44	1.00									
11	0.54	0.69	0.67	0.74	0.57	0.40	0.50	0.48	0.54	0.50	1.00								
12	0.72	0.64	0.30	0.31	0.30	0.53	0.32	0.50	0.38	0.53	0.26	1.00							
13	0.75	0.58	0.42	0.48	0.42	0.44	0.44	0.74	0.60	0.67	0.36	0.76	1.00						
14	0.69	0.54	0.09	0.52	0.29	0.50	0.20	0.38	0.27	0.40	0.33	0.61	0.54	1.00					
15	0.67	0.58	0.63	0.48	0.53	0.56	0.56	0.53	0.60	0.44	0.36	0.67	0.60	0.36	1.00				
16	0.38	0.48	0.50	0.45	0.50	0.40	0.80	0.37	0.47	0.67	0.53	0.22	0.35	0.21	0.35	1.00			
17	0.72	0.72	0.60	0.54	0.50	0.53	0.63	0.80	0.67	0.63	0.52	0.76	0.35	0.57	0.67	1.00			
18	0.45	0.54	0.59	0.35	0.59	0.37	0.75	0.47	0.44	0.75	0.60	0.32	0.44	0.20	0.33	0.80	0.74	1.00	
19	0.58	0.42	0.63	0.64	0.74	0.33	0.89	0.63	0.70	0.67	0.54	0.29	0.50	0.36	0.50	0.71	0.67	0.67	1.00

probably reflects the outcrossing nature of walnut since similar results have been obtained with RAPDs in other outcrossing fruit and nut tree species such as pistachio (Hormaza et al., 1994) or olive (Fabbri et al., 1995).

As expected, most of the fragments amplified from DNA obtained from the progenies were also present in the parents. Nevertheless, three markers (K01-445, G16-700 and R15-490) were, in some cases, present in the progeny and absent in both parents. Thus, K01-445 was present in 'Vina' and absent in 'Payne' and 'Franquette'; G16-700 was present in 'Sunland' and absent in PI-159568 and 'Lompoc'; R15-490 was present in 'Serr' and absent in 'Lompoc' and 'PI-159568'. The occurrence of non-parental bands has been reported in previous studies with RAPDs (Hunt & Page, 1992; Riedy et al., 1992; Aruna et al., 1993; Ayliffe et al., 1994; Pooler & Scorza, 1995) and different explanations have been suggested, such as formation of heteroduplex molecules between alternate RAPD alleles, mutations or recombination events within the primer binding sites or inside the amplified fragments, competition for primer binding sites or somatic rearrangements in perennial plants.

The similarity values based on 23 RAPDs (Table 3) ranged from 0.09 for 'Cisco' and 'Sharkey' to 0.89 for 'Lompoc' and 'Waterloo'. UPGMA cluster analysis of the similarity matrix (Figure 3) separated the walnut genotypes included in this study into two groups whose differences were basically related to the original sources of the genotypes used as parents in the breeding program (Figure 1). The first group comprises 'Sharkey' and 'PI-159568', two genotypes originating from Asia, and some of their progeny: '56-224', 'Chandler', 'Howard', 'Serr' and 'Sunland'. The second group contains the Californian ('Marchetti', 'Payne', 'Waterloo') and French ('Conway Mayette', 'Franquette' and 'Meylan') parental cultivars and most of their progeny ('Cisco', 'Lompoc', 'Pedro', 'Tehama', 'Tulare' and 'Vina'). This is not unexpected since the three French parental cultivars probably originated in the same French region (Isere). In general, the cultivars sharing common parents tend to group together and with at least one of the parents.

Among all the genotypes tested, 'Sharkey' appears to be the most distantly related to all the others except to its progeny ('56-224') and to 'PI-159568' with a similarity value of 0.69 and 0.61 respectively. 'PI-159568' itself shows a high similarity value only with its progeny, and with 'Sharkey' and 'Sharkey's progeny ('56-224'). The markers G18-990, R19-755 and T06-400 are only present in 'PI-159568' and 'Sharkey' and some of their progeny and the marker R14-1130 is only absent in 'PI-159568' and 'Sharkey'. It is interesting to note that those two genotypes are the only ones

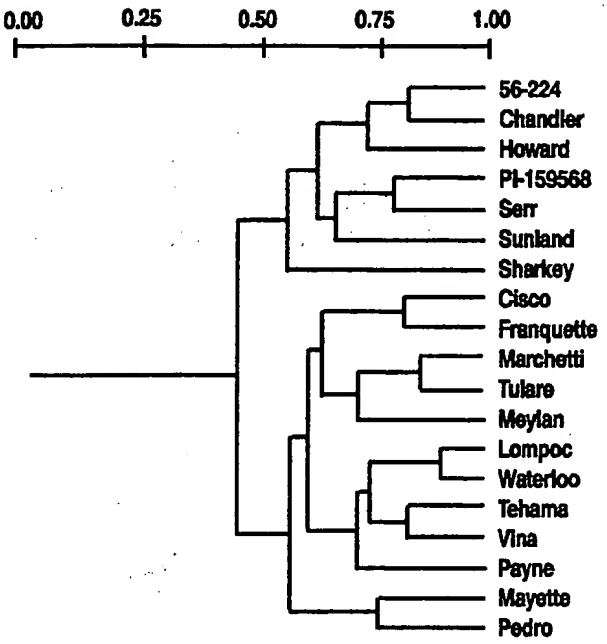


Figure 3. Dendrogram of the 19 walnut genotypes studied generated by UPGMA cluster analysis of the similarity values shown in Table 3.

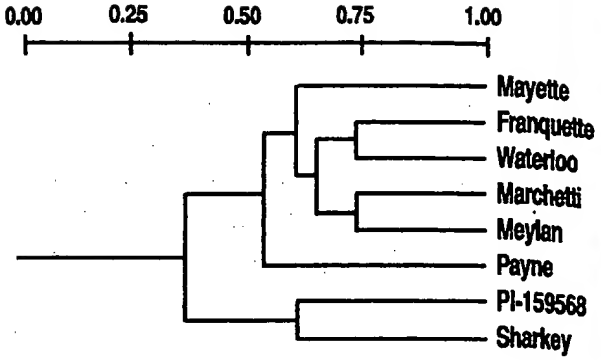


Figure 4. Dendrogram of the parental cultivars generated by UPGMA cluster analysis of the similarity values shown in Table 3.

in this study originating from Asia: i.e., 'PI-159568' from Afghanistan and 'Sharkey' probably from China (Tulecke & McGranahan, 1994).

The cultivar Payne, ancestor of most of the cultivars tested, showed, as expected, a similarity value of 0.50 or more with all 10 cultivars related to it except 'Cisco' (similarity value of 0.47) and 'Sunland' (similarity value of 0.44), two cultivars that are two generations away from 'Payne' in the pedigree.

In our research, as detected with other crop species (Aruna et al., 1993; Dunemann et al., 1994; Dweikat et al., 1993; Hallden et al., 1994), we observe a fairly close relationship between the known pedigree and the genetic similarity obtained with RAPDs. This is of great interest in breeding tree crop species since very often the pedigree of the cultivars is unknown. However, it was not possible to compare the genetic similarity values estimated with the Nei & Li index with the coefficients of coancestry (Falconer, 1989) due to three main reasons. One, that the available pedigree consists of a maximum of just two generations. Second, the French and Californian cultivars are probably related since the genetic base of the Californian cultivars has been enriched with introductions from France. Third, although the markers we have used allow us to unequivocally distinguish all the cultivars studied, a higher number of markers will probably be needed to obtain a dendrogram that accurately reflects the similarity matrix; in our case the correlation coefficient between the cophenetic matrix developed from the dendrogram and the similarity matrix was 0.65. However, if only the eight initial parental genotypes are analyzed, the dendrogram obtained (Figure 4) is a very good representation of the similarity matrix since the correlation coefficient between the two matrixes is 0.86. Moreover, in this case, two clear clusters are obtained: one with the genotypes of Asian origin, 'PI-159568' and 'Sharkey', and the other with the French and Californian cultivars. This is the expected result based on the fact that the gene pool of the Californian cultivars has been enriched with French germplasm. Comparisons of genetic distances obtained with molecular markers and theoretical data based on pedigree information have been already made in different herbaceous species and generally molecular marker-based measures of genetic distance agree with pedigree information (Dudley, 1994). Since pedigree and passport data are often unknown or incomplete for many fruit and nut tree species (Warburton & Bliss, 1996) RAPDs can be a useful tool to assess the degree of similarity of accessions or cultivars in these woody species in order to select the best parents to obtain new genetic combinations; this is especially important if we consider the long generation times of most fruit and nut tree species and, consequently, the length of the breeding process.

The results obtained show first that the RAPD technique can detect enough polymorphism to differentiate among walnut genotypes, even among cultivars closely related because of their common parents (for example 'Lompoc'-'Tehama' and 'Chandler'-'Howard'). Second, a general pattern of separation between Californian-European and Asian genotypes was obtained in this study confirming the results previ-

ously reported with RFLPs by Fjellstrom et al. (1994). Third, the RAPD method is a relatively simple technique to study genetic relationships in walnut thus allowing the study of the influence of genetic drift and selection which cannot be predicted using pedigree information alone. From the data obtained in this study we can conclude that RAPD technology can be useful in current walnut breeding programs, allowing the identification of new cultivars as well as the assessment of the genetic similarity among different genotypes.

Acknowledgements

J.I. Hormaza was supported by an Instituto Nacional de Investigaciones Agrarias (INIA) fellowship from the Spanish Ministry of Agriculture. F.P. Nicese was supported by a fellowship from Consiglio Nazionale delle Ricerche, Italy. We gratefully acknowledge S. Arulsekar, V.S. Polito and K. Woeste for their helpful comments on the manuscript.

References

Aleta, N., C. Olarte, M.J. Truco & P. Arus, 1989. Identification of walnut cultivars by isozyme analysis. Acta Horticulturae 284: 91–96.

Aleta, N., M. Rovira, A. Ninot & P. Arus, 1993. Inheritance of four isozymes in walnut. Acta Horticulturae 311: 62–65.

Aly, A.M.M., R.G. Fjellstrom, G.H. McGranahan & D.E. Parfitt, 1992. Origin of walnut somatic embryos determined by RFLP and isozyme analysis. HortScience 27: 61–63.

Arulsekar, S., D.E. Parfitt & G.H. McGranahan, 1985. Isozyme gene markers in *Juglans* species. J Heredity 76: 103–106.

Arulsekar, S., G.H. McGranahan & D.E. Parfitt, 1986. Inheritance of phospho-glucomutase and esterase isozymes in Persian walnut. J Heredity 77: 220–221.

Aruna, M., P. Ozias-Akins & M.E. Austin, 1993. Genetic relatedness among rabbiteye blueberry (*Vaccinium ashei*) cultivars determined by DNA amplification using single primers of arbitrary sequence. Genome 36: 971–977.

Ayliffe, M.A., G.J. Lawrence, J.G. Ellis & A.J. Pryor, 1994. Heteroduplex molecules formed between allelic sequences cause nonparetnal RAPD bands. Nucl Acids Res 22(9): 1632–1636.

Bachmann, K., 1994. Molecular markers in plant ecology. The New Phytologist 126: 403–418.

Doyle, J.J. & J.L. Doyle, 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19: 11–15.

Dudley, J.W., 1994. Comparison of genetic distance estimators using molecular marker data. Analysis of molecular marker data. Joint plant breeding Symposia series. Am Soc Hort Sci Crop Sci Soc USA, pp. 3–7.

Dunemann, F., R. Kahnau & H. Schmidt, 1994. Genetic relationships in *Malus* evaluated by RAPD fingerprinting of cultivars and wild species. Plant Breed 113(2): 150–159.

Dweikat, I., S. Mackenzie, L. Morris & H. Ohm, 1993. Pedigree assessment using RAPD-DGGE in cereal crop species. Theor Appl Genet 85: 497–505.

Fabbri, A., J.I. Hormaza & V.S. Polito, 1995. Random amplified polymorphic DNA analysis of olive (*Olea europaea* L.) cultivars. J Am Soc Hort Sci 120(3): 538–542.

Falconer, D.S., 1989. Introduction to Quantitative Genetics. 3rd. ed. Longman, London.

Fjellstrom, R.G. & D.E. Parfitt, 1994a. RFLP inheritance and linkage in walnut. Theor Appl Genet 89(6): 665–670.

Fjellstrom, R.G. & D.E. Parfitt, 1994b. Walnut (*Juglans* spp.) genetic diversity determined by restriction fragment length polymorphisms. Genome 37(4): 690–700.

Fjellstrom, R.G. & D.E. Parfitt, 1995. Phylogenetic analysis and evolution of the genus *Juglans* (Juglandaceae) as determined from nuclear genome RFLPs. Plant Syst Evol 197(1–4): 19–32.

Fjellstrom, R.G., D.E. Parfitt & G.H. McGranahan, 1994. Genetic relationship and characterization of Persian walnut (*Juglans regia* L.) cultivars using restriction fragment length polymorphism (RFLPs). J Am Soc Hort Sci 119(4): 833–839.

Forde, H.I. & G.H. McGranahan, 1996. Walnuts. In: J. Janick & J.N. Moore (Eds.), Fruit Breeding, Vol III. Nuts, pp. 241–273. John Wiley & Sons, New York.

Germain, E., I. Hanquier & R. Monet, 1993. Identification of eight *Juglans* spp. and their interspecific hybrids by isoenzymatic electrophoresis. Acta Horticulturae 311: 73–81.

Hallden, C., N.O. Nilsson, I.M. Rading & T. Sall, 1994. Evaluation of RFLP and RAPD markers in a comparison of *Brassica napus* breeding lines. Theor Appl Genet 88(1): 123–128.

Hormaza, J.I., L. Dollo & V.S. Polito, 1994. Determination of relatedness and geographical movement of *Pistacia vera* L. (pistachio; Anacardiaceae) germplasm by RAPD analysis. Econ Bot 48(4): 349–358.

Hunt, G.J. & R.E. Page, 1992. Patterns of inheritance with RAPD molecular markers reveal novel types of polymorphism in the honey bee. Theor Appl Genet 85(1): 15–20.

Lamboy, W.F., 1994. Computing genetic similarity coefficients from RAPD data: the effects of PCR artifacts. PCR Methods Appl 4: 31–37.

Louskas, M., K. Pontikis, I. Vergini & K. Papalexandras, 1984. Identification of walnut and filbert cultivars by isoenzymatic electrophoresis. Acta Horticulturae 311: 73–81.

Molvolti, M.E., M. Pigliucci, F. Cannata & S. Fineschi, 1993. Genetic variation in Italian populations of *Juglans regia*. Acta Horticulturae 311: 86–91.

Malvolti, M.E., S. Fineschi & M. Pigliucci, 1994. Morphological integration and genetic variability in *Juglans regia* L. J Heredity 85: 389–394.

McGranahan, G.H. & C. Leslie, 1990. Walnuts (*Juglans*). Acta Horticulturae 290: 905–951.

McGranahan, G.H., W. Tulecke, S. Arulsekar & J.J. Hansen, 1986. Intergeneric hybridization in the Juglandaceae: *Pterocarya* spp. × *Juglans regia*. J Am Soc Hort Sci 111(4): 627–630.

Nei, M. & W.H. Li, 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci USA 76: 5269–5273.

Newbury, H.J. & B.V. Ford-Lloyd, 1993. The use of RAPD for assessing variation in plants. Plant Growth Regulation 12: 43–51.

Pooler, M.R. & R. Scorza, 1995. Aberrant transmission of RAPD markers in haploids, doubled haploids, and F_1 hybrids of peach: observations and speculations on causes. Scientia Horticulturae 64: 233–241.

Riedy, M.F., W.J. Hamilton & C.F. Aquadro, 1992. Excess of non-parental bands in offspring from known primate pedigress assayed using RAPD PCR. Nucl Acids Res 20(4): 918.

Rieseberg, L.H., 1996. Homology among RAPD fragments in interspecific comparisons. Mol Ecol 5: 99–105.

Rink, G., G. Zhang, Z. Jinghua, F.H. Kung & E.R. Carroll, 1994. Mating parameters in *Juglans nigra* L. seed orchard similar to natural population estimates. Silvae Genetica 43(4): 261–263.

Serr, E.F., 1969. Persian walnuts in the western states. In: R.A. Jaynes (Ed.), Handboof of North American Nut Trees, pp. 240–263. W.F. Humphry Press Inc., Geneva, New York.

Solar, A., J. Smole & F. Stampar, 1993. Identification of walnut cultivars by pollen isozymes. Acta Horticulturae 311: 95–99.

Solar, A., J. Smole, F. Stampar & M. Virscekmarn, 1994. Characterization of isozyme variation in walnut (*Juglans regia* L.). Euphytica 77(1–2): 105–112.

Tulecke, W. & G.H. McGranahan, 1994. The walnut germplasm collection at the University of California, Davis; a description of the collections and a history of the breeding program of Eugene F. Serr and H.I. Forde. Genetic Resources Conservation Program, University of California, Davis, Rpt. 13.

Warburton, M.L. & F.A. Bliss, 1996. Genetic diversity in peach (*Prunus persica* L. Batch) revealed by randomly amplified polymorphic DNA (RAPD) markers and compared to inbreeding coefficients. J Am Soc Hort Sci 121(6): 1012–1019.

Weising, K., H. Nybom, K. Wolff & W. Meyer, 1995. DNA fingerprinting in plants and fungi. CRC Press Inc., Boca Raton, Florida.

Wenheng, C.S.Y., 1984. Taxonomic studies of ten species of the genus *Juglans* based on isozymic zymograms. Acta Hort Sinica 14(2): 90–96.

Williams, J.G.K., A.R. Kubelik, K.J. Livak, J.A. Rafalski & S.V. Tingey, 1990. DNA polymorphism amplified by arbitrary primers are useful as genetic markers. Nucl Acid Res 18: 6531–6535.

Williams, J.G.K., M.K. Hanafey, J.A. Rafalski & S.V. Tingey, 1993. Genetic analysis using Random Amplified Polymorphic DNA markers. Methods Enzymol 218: 704–740.

Woeste, K., G.H. McGranahan & R. Bernatzky, 1996a. Randomly amplified polymorphic DNA loci from a walnut backcross [(*Juglans hindsii* × *J. regia*) × *J. regia*]. J Am Soc Hort Sci 121: 358–361.

Woeste, K., G.H. McGranahan & R. Bernatzky, 1996b. The identification and characterization of a genetic marker linked to hypersensitivity to the cherry leafroll virus in walnut. Mol Breed 2(3): 261–266.